Vol. 29, No. 4 August, 1945

Journal of Applied Psychology

EDITED BY: DONALD G. PATERSON, UNIVERSITY OF MINNESOTA

Consulting Editors

PAUL S. ACHILLES, *Psychological Corporation;* WALTER V. BINGHAM, *A.G.O., War Department;* HAROLD E. BURTT, *Ohio State University;* ARTHUR I. GATES, *T. C. Columbia University;* JOHN G. JENKINS, *University of Maryland;* IRVING LORGE, *T. C. Columbia University;* QUINN MCNEMAR, *Stanford University;* WILLARD C. OLSON, *University of Michigan;* JAMES P. PORTER, *Ohio University;* EDWARD K. STRONG, JR., *Stanford University;* MORRIS S. VITELES, *University of Pennsylvania;* JOSEPH ZUBIN, *N. Y. Psychiatric Institute.*

Table of Contents

Intelligence and Adjustment Measurements in the Selection of Radio Tube Mounters: G. FORLANO AND F. H. KIRKPATRICK 257

Single-Item Tests for Psychometric Screening: LIEUT. H. M. HILDRETH 262

Range of Interests: LIEUT. R. F. BERDIE 268

A Note on the Problem of Brain Damage in Rehabilitation and Personnel Work: H. F. HUNT 282

Personality Patterns of Adolescent Girls: II. Delinquents and Non-Delinquents: D. F. CAPWELL 289

The Construction of the Guilford-Martin Inventory of Factors G-A-M-I-N: H. G. MARTIN 298

Measuring Progress in Radio Training: G. L. MACDONALD 301

A Study of the Effect of Music Distraction on Reading Efficiency: M. T. HENDERSON, A. CREWS, AND J. BARLOW 313

Book Reviews 318

New Books, Monographs, and Pamphlets 321

Published Bi-monthly by The American Psychological Association, Inc.
With the Cooperation of The American Association for Applied Psychology
Prince and Lemon Sts., Lancaster, Pa., and Northwestern University, Evanston, Illinois

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the Act of March 3, 1879
Copyright 1945, by The American Psychological Association, Inc.

Journal of Applied Psychology

Vol. 29, No. 4 August, 1945

Intelligence and Adjustment Measurements in the Selection of Radio Tube Mounters

George Forlano * and Forrest H. Kirkpatrick

Radio Corporation of America *Bethany College*

The general problem with which this study was concerned is the increase of worker efficiency in radio tube mounting jobs. Specifically the problem reduced to: will intelligence and adjustment tests be effective in bringing about this desired increase in efficiency? There have been many claims as to the effectiveness of adjustment tests in particular in the selection of more desirable workers, but there have been too few quantitative presentations of proof. Vague opinions of personnel officers that the use of tests brings superior employees are too often accepted without evidence.

The subjects employed in this experiment were twenty female tube mounters in one department of the Indianapolis Plant of the Radio Corporation of America. Tube mounters assemble the very small tube elements to the bases of radio tubes. It is very close work requiring considerable finger and hand dexterity. Several mounters, each performing distinct operations, work together on each tube. The subjects were new employees who had taken the required placement tests, and had been assigned as tube mounters. New employees were used in the experiment because it was desirable to keep all regular workers on the production line.

The tests used for the purposes of this experiment were the (1) Otis Self-Administering Test of Mental Ability Form B, (2) the Bell Adjustment Inventory, adult form, and (3) the Washburne Social Adjustment Inventory. Of the latter two tests only the scores which are designed to measure social adjustment were used. This included the "social" score of the Bell test and the "alienation" score of the Washburne test. These were administered before the employees started work. The subjects also took the regular aptitude battery (a vision test, a two-hand coordination test, and manual dexterity tests) with which we shall not concern ourselves in this report.

* Now in the U. S. Army.

The criterion used for this experiment was ratings by the supervisor in charge of the group. Each worker was rated for the extent to which she took hold of the job and performed efficiently. These ratings were given on the basis of one month's observation of the worker at her task. The supervisor's ratings were in the following terms: G, indicating a good employee; and F, indicating a fair employee.

For purposes of examination of the data, test score distributions were divided into three groups and scores within the groups assigned the values: above average, average, or below average. Table 1 gives the groupings

Table 1

Distribution of Test Score Values to Various Groupings

Test	Grouping		
	Above Average	Average	Below Average
Otis Mental Ability (I.Q.)	115 and over	95 to 114	94 and below
Bell Social Scores	5 to 8	9 to 19	20 to 24
Washburne Alienation Scores	0 to 5	6 to 15	16 and above

Table 2

Showing Mounters' Tests and Related Data

Tube Mounter	Supervisor's Ratings	Otis I.Q.'s	Personality Rating	Composite Intelligence and Personality
1.	G	125	+av.	12
2.	F	122	−av.	6
3.	G	122	av.	9
4.	G	122	+av.	12
5.	F	118	−av.	6
6.	G	118	av.	9
7.	F	118	av.	9
8.	F	114	−av.	3
9.	G	111	+av.	9
10.	F	101	−av.	3
11.	G	98	av.	6
12.	G	97	+av.	9
13.	F	95	av.	6
14.	F	91	av.	3
15.	F	80	−av.	0
16.	F	76	−av.	0
17.	F	76	av.	3
18.	F	72	−av.	0
19.	G	69	av.	3
20.	F	95	av.	3

used. These divisions were based on distribution of scores into upper, middle, and lower thirds.

The two personality test scores were then combined in such a way that if a worker made above average on both tests she was rated "+ average" for personality; if she received above average on one test and below on the other, or if she was average on both, she was rated "average" in personality. Likewise, if the worker made below average on both tests she was rated "−average" in personality. These are the figures shown in Table 2 under the heading "Personality Rating."

Table 2 also shows a column headed "Composite Intelligence and Personality." These figures were obtained by weighting an above average personality rating or Otis score as 6, an average rating or score as 3, and a below average rating or score as 0, and then adding the two weights for each person. Hence, the composite could range from 0 (below average Otis score + below average personality rating) to 12 (above average Otis score + above average personality rating). Table 2, then, contains the basic figures from which experimental results were studied.

Intelligence and Job Success

Table 3 shows the relationship of intelligence test scores to the ratings of job success. For employees with average and above average scores there were exactly the same number rated "good" by supervisors as there

Table 3

Distributions of Intelligence Test Scores for "Good" and "Fair" Employees

Test Scores	Rated "Good"	Rated "Fair"
Above Average	3	3
Average	3	3
Below Average	1	6

were rated "fair." However, 6 out of 7 who were below average in intelligence had been rated as "fair" by the supervisor. It would seem from these results that low intelligence would go with poor job success while above average intelligence does not give advantage over average. This might have been expected since, although it takes a certain amount of intelligence to learn the job, there is not much to tax mental capacity after it is learned.

Social Adjustment and Job Success

Table 4 summarizes the relationship between personality test scores for social adjustment and ratings of job success. The scale indicating "social adjustment" here is the combination of Bell and Washburne scores

given under "Personality Rating" in Table 2. These results are rather striking in that all of the people scoring above average on the tests also were rated "good" by their supervisor, while all of the people scoring below average on the tests were rated "fair" by the supervisors. Whether this

Table 4

Distributions of Personality Ratings for "Good" and "Fair" Employees

Test Scores	Rated "Good"	Rated "Fair"
Above Average	4	0
Average	4	5
Below Average	0	7

indicates that a socially adjusted personality is an important factor in doing good work, or that the supervisor was more influenced by the girls' "personalities" than their performance, is uncertain. Certainly the tests were measuring something which was related to the supervisor's notion of good and fair employee risks.

Job Success and the Composite Score

When the intelligence and personality scores are combined in the manner indicated in Table 1, distinction between "good" and "fair" employees is apparent. Table 5 shows that by eliminating all the employees

Table 5

Composite Intelligence and Personality Score Distributions for "Good" and "Fair" Workers

Composite Score	Rated "Good"	Rated "Fair"
12	2	0
9	4	1
6	1	4
3	1	5
0	0	3

who had a composite score of 3 or less, we would have left 88% of the "good" employees and only 39% of the "fair" employees.

Moreover, by eliminating employees whose composite scores are 6 or less, all but one, or 92%, of "fair" employees would be eliminated. At the same time only 2, or 25%, of "good" employees would be lost. It would seem from these figures that, in times of a favorable labor market, these tests would be very effective in selecting workers able to meet with the approval of supervisors.

Summary

This study was concerned with the effectiveness of certain personality and intelligence measures in predicting the job success of 20 tube mounters. Relationships were shown between supervisory ratings of job success and scores on the Otis Self-Administering Test of Mental Ability, Form B, the social scale of the Bell Adjustment Inventory, and the alienation scale of the Washburne Social Adjustment Inventory. The following conclusions were drawn:

(1) Low intelligence scores tended to indicate the poorer workers. Average or above average scores did not discriminate between "good" and "fair" workers.

(2) All workers scoring below average in social adjustment had been rated only "fair" by the supervisor while all workers scoring above average in social adjustment had been rated "good" by the supervisor. The supervisor had no previous knowledge of test scores.

(3) A composite of intelligence and personality scores was shown to be effective in predicting the subsequent success of new tube mounters.

Received July 28, 1944.

Single-Item Tests for Psychometric Screening *

Lieut. H. M. Hildreth, USNR

A series of 10 Single-Item Tests are presented in this article. These tests, designed for psychometric screening, were developed at a Naval training station in response to the need for rapid methods of examining the intelligence of large numbers of recruits. They take from 1 to 2 minutes per man to administer, and do not require optimal testing conditions. Standardized on 1,500 cases they have also been tested in actual practice and found to be practicable. In a 3-month trial period, during which the tests were used as a basis for accepting recruits for service, the error involved was found to be less than 1/10 of 1 per cent.

A New Approach to the Problem of Screening

Single-Item Tests are the result of a new approach to the screening problem of increasing speed of testing without losing accuracy. The usual approach has consisted of abbreviating, and devising short forms of, standard mental tests. These short forms bring testing time down from an hour to around 10 or 15 minutes, without appreciable loss of accuracy; but this is about the limit to which standard-type tests can be skeletonized without sacrificing reliability.

The reason for the limitation of this approach lies in the psychometric requirements imposed by the purpose of mental measurement. In the past this purpose has always been to determine the true or *maximum* mental ability of the individual. Standard tests, including those recently streamlined for screening, have always been designed and constructed to measure this maximum.

In psychometric screening the individual's maximum is not a matter of concern. The purpose in screening is to make sure that a recruit's mental ability is not below a given *minimum* and that he is thereby acceptable for naval or military service. Tests for acceptance need measure only this minimum; it is not necessary to measure maximum ability.

Recognition of this basic difference in purpose makes it possible to dispense with 3 traditional psychometric requirements. In tests for maximum ability: (1) many items are ordinarily included in the test so that the individual will have ample opportunity to demonstrate his full

* The opinions or assertions contained in this article are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

ability; (2) failure is accorded as much significance as success, since the failure level of an individual must be determined if his maximum performance is to be known; and (3) scoring of failure requires in turn that maximum-ability tests be administered under optimal testing conditions. Unlike success, which demonstrates the presence of ability, failure does not necessarily indicate lack of ability. It may be due to distraction, confusion or anxiety, and until these and similar extraneous causes are eliminated, failure cannot be taken to mean "lack of ability."

In testing for minimal mental ability none of these requirements need be considered. One success is sufficient to establish a minimum. A recruit who is successful on a single item demonstrates he has at least the degree of mental ability called for by that item. He may well have more, but it is certain he does not have less. If the measuring power of the particular item has been determined, then a minimal mental ability can be ascribed to the man at once and without further testing.[1]

Ordinarily, however, only the average difficulty of test items is known. The minimal mental ability which they represent is not known. To obtain this information, and make possible the use of single items as screen tests, the special scoring technique described below was devised.

Scoring Technique

The technique used in scoring items for screening consists of tabulating for a normal group the mental ages of all those who are successful on a given item, and then determining the point at which this distribution becomes asymptotic to the mental-age base line. Those items found to represent a minimal mental ability of the desired degree are suitable as screen tests.

The desired minimum in the present case was 11 years, or 132 months, mental age, the minimal mental ability for Naval service originally set by official directives. Thirty items from well-known tests (1, 2, 3) were selected for scoring. In conjunction with a Stanford-Binet examination these 30 items were administered under optimal conditions to 1,500 men of military age. Care was taken to assure a normal distribution in respect to mental age, particularly in the lower brackets.

Each of the 30 experimental items was then scored for minimal mental ability. (1) A frequency distribution was made of the mental ages of all those who passed the given item. Failures were disregarded. (2) The point at which the distribution approached zero was located. For practi-

[1] The foregoing applies to free-response items such as those used in the S-I Tests. Success cannot be considered prima facie evidence of ability in multiple-choice or true-false items where an appreciable element of chance is present. Coaching, of course, invalidates any mental test.

cal purposes this point was set at the first percentile, with the one-half percentile point also being computed. (3) The mental ages corresponding to these points were determined. If the mental age at the first percentile was 132 months or greater, the item was considered suitable as a Single-Item Test for screening.

In the "Squares" problem below, for example, all individuals in the standardization group passing the item were separated out and their mental ages were tabulated in a frequency distribution. The first percentile of this distribution was computed and found to be 132 months. Since this value was not below the desired minimum, the item was considered suitable as S-I Test. The chances are less than 1 in 100 that a recruit who passes this test during screening will have a mental age below 132 months.

Of the 30 experimental items scored in this manner 10 were found to guarantee mental ability above the Navy minimum. These 10, designated as Single-Item Tests, are described below.[2]

Description of Single-Item Tests

In the column headed $P = .990$ is listed (in months) the mental age value at the first percentile. The probability is .99 that an individual passing the S-I Test will have a mental age at or above that listed in the table. For column $P = .995$ the chances are 199 in 200 that the individual will not be below the mental age listed there. These probabilities are in terms of a normally distributed male group of military age, as found at an induction center. This is the group customarily dealt with in screening.[3]

Scoring of the Kent and Stanford-Binet items is in most instances identical with that originally described for these items.

1. Question:	"How much is:	$P = .990$	$P = .995$
7 times 7	10 times 10	132	128
8 times 8	11 times 11		
9 times 9	12 times 12		

Scoring: Pass if subject makes no more than two errors.

[2] It will be noted that some of the values in this table are below 132 months, but not below 126. It has been found in practice that 126 is a safe value to use because of the conditions under which screening is usually done, and because Stanford-Binet scores on subnormals tend to run somewhat lower than the Wechsler-Bellevue (4) which is frequently used for determining mental age. Scoring of Wechsler-Bellevue items for screening is now under way.

[3] When testing is limited to doubtful cases all of whom are suspected of mental deficiency success may be conservatively interpreted in terms of this group. The probability is less than .05 that an individual whose mental age is below 11 years will pass any given S-I Test. The null-type hypothesis that the individual is defective is then disproved by his success.

2. Question: "What does the word PRICELESS mean?	133	126
Scoring: Full credit for such responses as "very valuable," "you can't buy it for any amount of money," "it is worth a lot." If the subject says "It has no price," ask what he means. Make sure he really knows the meaning of the word.		
3. Question: "If your shadow points toward the northeast, in which direction is the sun?"	134	128
Scoring: Credit if response is "Southwest."		
4. Question: "What does the word TOLERATE mean?"	142	139
Scoring: Full credit for such responses as "it means to put up with things," "it means you stand for what people do without getting mad," "it means to be lenient," "to endure." Do not credit such responses as "I tolerate you" unless the subject can follow with a clear explanation. Do not credit such responses as "it's a sickness—the cholery."		
5. Question: "What is the next number?" 1 2 4 8 16 —	139	137
These numbers are presented in written form.		
If subject hesitates, say "These numbers go up in a certain order; what would be the next one?"		
Scoring: Credit only if the subject says "32" and is able to explain that the numbers are being doubled.		
6. Question: "Why does the moon look bigger than the stars?" "What time of day is your shadow shortest?"	136	132
Scoring: Full credit if both questions are answered correctly.		
For the first question credit such responses as "the moon is closer (nearer) to the earth," "the stars are farther away." No credit for such answers as "the moon is bigger"; give assurance that the moon is really smaller, and repeat the question.		
For the second question, credit "noon" or any time between 11:00 and 1:00. No credit for "night" or "4 o'clock."		
7. Question: "What is the opposite of WINTER?" "What is the opposite of WAR?" "What is the opposite of BROAD?" "What is the opposite of ALIKE?" "What is the opposite of DEEP?"	135	129
Scoring: Full credit if subject can give the opposite of four of these words.		
8. Question: "Make a good sentence out of these words." FOR THE STARTED AN WE COUNTRY EARLY AT HOUR	129	126
These are presented to the subject in written form.		
Scoring: Full credit for "we started for the country at an early hour," "at an early hour we started for the the country," "We started at an early hour for the country."		
9. Question: "What does the word MARS mean?" "What does the word HYSTERICS mean?" "What does the word BRUNETTE mean?"	130	127

Scoring: Pass the subject if the meaning of each of the words can be given.

For MARS give credit for "planet" or "the God of War." No credit for "it is up in the sky," "it is a place," "it is a country," "it is a candy bar."

For HYSTERICS give credit for "it is nervous sickness," "it is like having a fit of crying," "to get excited and go all to pieces." No credit for "it is about old times" (historic).

For BRUNETTE give credit for "dark complexion," "black hair," "a person who is dark." No credit for "a blonde," "it is a woman," "red hair."

10. Question: "In what way are a KNIFE–BLADE, a PENNY, and a PIECE OF WIRE alike?"	129	126

Scoring: Full credit if response is "metal." No credit for "copper," "steel," "they are useful," or "they're all hard."

	$P = .990$	$P = .995$

Use of S-I Tests in Screening

Each of the 10 S-I Tests above is an independent test of acceptance. Success on any one of the tests demonstrates mental ability above the minimum for Naval training.[4] No attention need be given to providing optimal testing conditions. A distracting situation does not invalidate an S-I Test; on the contrary it increases the significance of the recruit's success. When privacy is not available and the replies of one man can be overheard by those next in line, the tests can be used in rotation, each man receiving a different question.

The majority of men pass the first test given them promptly. Failure on any or all of the S-I Tests, however, does not mean that a recruit is necessarily below the required mental minimum. Confusion rather than lack of ability may cause the failure. If the recruit is unsuccessful in several tries, he should be held over for further examination.

This subsequent examination may include more elaborate phases of acceptance testing (5). It may also involve testing for rejection with the usual full-length psychological tests. These "many-item" tests, designed to measure maximum intelligence, are necessary before a decision to reject can be made, and the more hours available for thorough rejection-testing of this group the more men can be salvaged for service. Single-Item Tests, by speeding up acceptance, save time for the more careful examination of these borderline cases.

[4] Present data indicate that when testing is continued until 2 S-I Tests are passed, 3 months may be added to each of the lower values. E.g. if the first and second tests are passed the new minimal values would be 135 and 129 for the 2 degrees of probability. In some cases, but not all, this will establish a higher minimum than that given for either of the 2 tests.

Summary

Ten Single-Item Tests for use in psychometric screening are described. These tests, the result of a new approach to the problem of screening, require 1 to 2 minutes per man to administer. Success on any one of them denotes mental ability above the minimum required for Naval training.

S-I Tests for rapid acceptance allow more time for rejection-testing of borderline cases, many of whom can be salvaged when additional examining time is available.

Received July 5, 1944.

References

1. Terman, L. M. *The measurement of intelligence.* New York: Houghton Mifflin Co., 1916.
2. Kent, G. H. *Oral test for emergency use in clinics.* Baltimore: The Williams & Wilkins Co., 1932.
3. Brown, M. A simple method for rapid estimation of intelligence in adults. *Amer. J. Orthopsychiat.*, 1942, 12, 411–414.
4. Wechsler, D. *The measurement of adult intelligence.* Baltimore: The Williams & Wilkins Co., 1941.
5. Hildreth, H. M., Wheeler, J. A., and Williams, S. B. *A psychometric procedure for screening mental defectives.* U. S. Naval Medical Bulletin (in press).

Range of Interests *

Ralph F. Berdie

Lieut. (jg), U.S.N.R.

Everyday observation and clinical experience suggest that the well adjusted individual has a wide range of interests and that other individuals who are experiencing conflicts within themselves and with their social environments tend to have more restricted ranges of interests. Narcissism accompanies many psychological malfunctions, and inability to attend to and maintain interest in occupational and extra-vocational activities often characterizes the neurotic adult. Extreme egocentricism is normal in infants and very small children. In physically mature individuals it indicates a lack of personality development or a regression resulting from various psychological frustrations. This regression may be as complete as that we observe in the advanced case of schizophrenia, or it may be more subtle.

Experience again suggests that the ranges of interests possessed by normal persons may vary widely and that some otherwise abnormal persons may have interests that do not differentiate them from normal people. A suspicion is justified that a few abnormal persons might have unusually wide ranges of interests and that this extensity of interest may or may not be related to their abnormality.

Very little relevant, systematic information is available regarding this problem of the relationship between the range of interests and other psychological attributes. The range of interests, as measured by the number of items marked as liked or disliked on the Strong Vocational Interest Blank, was found in one study [1] to be slightly correlated with high school and college grades and morale and social adjustment scores of the Minnesota Personality Scale. These correlations, ranging from $-.18$ to $+.23$ and determined on a sample of 411 college freshmen, were statistically significant but were too small to be of any value in predicting adjustment. The items liked on the Strong test are determined to a far greater extent by the person's vocational interest profile than by his personal adjustment. That is why the items were selected. The rela-

* The opinions or assertions contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

[1] Berdie, R. F. Likes, dislikes, and vocational interests. *J. appl. Psychol.*, 1943, 27, 180–189.

tionship found between range of interests and adjustment, however, suggests that items selected with the purpose of predicting adjustment rather than differentiating vocational interests might have a greater prognostic value.

The problem of predicting personal adjustment is particularly important to military psychologists attempting to eliminate those misfits from the service who, if allowed to remain, will eventually break down so as to endanger themselves and their fellows and shoulder the government with rehabilitation and pension costs.

Each man entering military service passes through at least one psychiatric screening. The type of screening depends on the available staff, the time allowed, and the facilities at hand, but in each case an attempt is made to determine if the man presents an adequate military risk or not. The pressure of work requires that the psychiatric screening be quick and even in those places where time for more extensive study is available, complete personality studies are not feasible.

Mental deficiency and illiteracy can be determined with relative simplicity. The frank psychotic, the organic deteriorate and the obvious misfit can be selected with but little more difficulty. Other cases, such as anxiety neurotics, hysterical personalities and constitutional psychopathic states, are more difficult to identify.

The problem attacked in this study concerned the usefulness of the range of interests in selecting potential psychological misfits at the beginning of military training. Does the range of interests possessed by an individual bear a relationship to his military adjustment and can this range be adequately measured? Are criteria available against which we can evaluate the predictive value of such a measure and finally, is such a measure useful in the military situation?

The study is divided into two sections. The first concerns a list of interest items presented orally to recruits. Here are discussed the questions of how stable the likes are from group to group, how the range of interests is related to ability and how well it differentiates between normal and abnormal recruits.

The second section of the study concerns a list of interest items presented in printed form to recruits. Here are discussed the questions of how reliably this list measures the range of interests, what the relationship between the range of interests, age, and education is, and how well it differentiates between normal and various groups of abnormal recruits.

Orally Presented List

A group of 31 activity interest items were pre-tested on a group of marine recruits and 22 of these items were liked by over 50 per cent of the

men. These 22 items were retained to form the interest scale. The items are presented in Figure 1.

Name———————————Age———Platoon No.———————

Highest school grade completed:————at———— years of age.

DIRECTIONS: Below is a list of things people do for recreation. Place a check (X) after the name of each thing you *LIKE TO DO*. After those things you dislike or have never done, leave a blank space.

Play checkers	———	Fishing	———
Play pool	———	Hiking	———
Play horseshoes	———	Go to carnivals	———
Dance	———	Go on picnics	———
Go on dates	———	Reading	———
Box	———	Go to movies	———
Play cards	———	Listen to radio	———
Go to parties	———	Roller skate	———
Play basketball	———	Watch baseball	———
Play baseball	———	Watch football	———
Play football	———	Go swimming	———

FIG. 1. Interest list presented to marine recruits immediately after physical examination.

The items were then presented in individual interviews to 200 marine recruits selected at random at the completion of their physical examination.[2] In the interview the examiner said,

> "I would like to find out what things you like to do. I will name these things to you and you tell me if you like to do them or not."

The items were presented in the order they appear in Figure 1, the examiner saying, "Do you like to play checkers?", "Do you like to go out on dates?", etc. With the last item, if the man said he was married, he was asked, "Did you like to go out on dates before you were married?" If the recruit said he liked the activity or disliked it, the examiner recorded the response. If he said he did not know or had no contact with the activity, this was recorded as an "indifference" response. Some subjective judgments were necessary in recording the responses but any statement connotating acceptance ("I like it sometimes") was recorded as a "like."

The list of items was then presented to 68 marine recruits recommended by the psychiatric unit to the aptitude board for inaptitude discharges. These recruits were all literate and none were mental defectives.

[2] The representativeness of this group is evidenced by the mean General Classification Test score of 101.8, SD = 16.3 for 172 of the men for whom G.C.T. scores were available.

Their difficulties included anxiety neuroses, hysterical states, epilepsy, enuresis, constitutional psychopathic states, post-traumatic syndromes, and psychasthenia. They were a very heterogeneous group of non-psychotic psychiatric cases.

Each man had been interviewed at least twice by a psychiatrist and twice by a psychologist and psychological test scores and verified social histories were available for many. In each case the decision had been made that their problems would prevent them from making an adequate military adjustment. These decisions were all made before the presentation of the interest list.

Two indices of range of interest were obtained from the responses to the items. The first consisted only of the number of items liked. The second was based upon weights assigned to each item upon the basis of the percentage of the number of cases liking the item. A man not liking to listen to the radio, an activity liked by 98 per cent of men, is more of a deviate than one who doesn't like to dance, an activity liked by only 53 per cent of men. By weighting each response in direct proportion to the extent to which it indicated an individual was a deviate, we hoped to obtain a more differentiating score. The results showed that the scores based upon the weights gave no better differentiation than the scores consisting merely of the number of activities liked.

Results

The stability of interests from group to group is important. Will interest items rank themselves in the same order of popularity in different samples or do conditions affecting the popularity of different activities vary from group to group to such an extent that interest in these activities cannot serve as a basis for constructing a scale for general use? Are there items which have a well defined place in the recreational hierarchy of adult males in America?

The 22 items were ranked according to the proportion of the first 50 subjects who reported liking the items and then ranked according to the proportion of the second 50 subjects who liked the items. The rank order correlation was .89. The same thing was then done on the basis of the likes of the first and second one hundred subjects and again the rank order correlation was .89.

In almost any group of American men selected at random, between 90 and 100 per cent will report they like to go to the movies, about one-half will report they like to dance. As reported later, age and perhaps educational status help determine the individuals' interests, along with multitudinous other factors, but in considering groups drawn from the general population, these interests form themselves into a remarkably stable order.

General intelligence has frequently been mentioned as one of the factors determining interests. In order to see what relationship existed between the total number of likes reported by an individual and his intelligence, comparisons on the basis of the number of likes was made between groups receiving high, medium and low scores on the General Classification Test. Test scores were available for 172 of the 200 men selected at random. They were divided into three groups; 37 who had G.C.T. scores of 110 and above; 47 who had scores of 89 and below, and 88 who had scores of between 90 and 109 inclusive. The cumulative percentage distributions of the number of likes for the three groups are presented in the first three columns of Table 1.

The average person in the low ability group likes 18 plus items. The average person of the medium ability group likes 18 plus items. The

Table 1

Distributions of Number of Likes Obtained with Orally Presented List for Three Different Ability Groups, Group of Discharged Recruits and Normal Recruits

No. of Items Liked	G.C.T. = 110 and Above		G.C.T. = 89 and Below		G.C.T. = 90–109		Inaptitude Discharges		Normal Recruits	
	Freq	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
22	1	100	1	100	5	100			7	100
21			3	98	11	95			15	96
20	6	97	5	92	16	82			30	89
19	2	81	9	81	9	64	1	100	21	74
18	9	76	6	62	11	54			35	63
17	4	51	5	49	10	41	5	98	22	46
16	7	41	6	39	11	30	2	91	30	35
15	4	22	6	26	3	17	1	88	14	20
14	3	11	2	13	5	14	3	86	10	13
13					2	8	7	82	3	8
12	1	3	1	9	2	6	9	72	5	7
11							8	59		
10			1	6	1	3	9	47	3	4
9			2	4	1	2	4	34	3	2
8					1	1	5	28	2	1
7							4	20		
6							3	15		
5							2	10		
4							2	7		
3							2	4		
2							1	1		
Total	37		47		88		68		200	

average person of the high ability group likes 17 items. Inspection of the columns reveals little or no significant difference between the number of items liked by the three groups. This is particularly evident when the figures for the ability groups are compared with the figures for the discharged and normal groups, which will be discussed later. The conclusion seems warranted that, with this particular list of items, the range of interests bears little relation to general intelligence.

The primary question about the list of items is how well does it differentiate between people who are making a good adjustment to military training and people who are not. The answer is found in the last two columns of Table 1.

The column to the left presents the accumulative percentile ratings of the 68 inaptitude cases and the column at the right the same thing for the 200 normal cases. These figures are based on simply the number of likes reported by the individuals. The range for the inaptitude cases is from 2 to 19, with fifty per cent of the cases liking 10 or fewer items. For the normal cases, the range is from 8 to 22, with fifty per cent of the cases liking 18 or fewer items. Of the inaptitude cases, 20 per cent like fewer items than any of the normals while of the normals, 37 per cent like more items than do any of the inaptitude cases. By choosing a critical point at 13 items, we would eliminate only 8 per cent of the normals and 82 per cent of the inaptitude cases. Such a small degree of overlapping is almost unprecedented when psychological instruments are used. Little question can be raised that the normal and atypical groups studied here most certainly differ on the basis of the number of likes they report.

Printed Interest List. The preceding discussion concerns data collected in a brief, standardized interview. The collection of information in personal interviews has certain disadvantages, however, when dealing with large groups. As has already been mentioned, this method in a few cases requires the interviewer to make judgments as to whether the subject's response indicates liking or not. To facilitate the collection of information and to eliminate any variability associated with the interviewer, the interest list was presented as a check list. A copy of the list and instructions as presented to the subjects is presented in Figure 1.

The mimeographed interest list was presented to 792 marine recruits, all the members of eleven consecutive platoons, immediately upon the completion of their physical examination and just prior to their interviews with the psychiatrists. Of these 792 men, 23 were selected by the psychiatrists to return for further examination because of suspected illiteracy and/or mental deficiency. Another group of 25 recruits was selected to return because of suggestions of other psychiatric or psychological defects —psychoneurosis, epilepsy, etc.

The check list was also presented to 114 recruits who had been thoroughly studied by the psychiatrists and psychologists and who were being recommended to the aptitude board by the psychiatric unit for inaptitude discharge. These were almost all of the recruits not of defective or borderline intelligence who were recommended to the board over a certain period of time. They were given the list after the final decision had been made to recommend their discharge. All of these recruits had at least two interviews with a psychiatrist and one with a psychologist. Most of them had been seen twice by a psychologist. Many had several contacts with both psychiatrists and psychologists and many had been carefully observed in a special observation unit or carefully studied in the neuropsychiatric ward of a Naval hospital. Social histories available for most of the recruits substantiated the information which led to the recommendation for their discharge. This group of discharged individuals included recruits diagnosed as having anxiety neurosis; constitutional psychopathic state, emotional instability; constitutional psychopathic state, inadequate personality; psychoneurosis, mixed; post-traumatic syndrome; epilepsey; enuresis; psychoneurosis, hysteria; constitutional psychopathic state, schizoid personality; hypochondriasis; psychoneurosis, psychasthenia; somnambulism; constitutional psychopathic inferior without psychosis; psychoneurosis, neurasthenia; cryptic nostalgia.

Results for the Printed Interest List. The range of scores obtained with the mimeographed list was greater and the distribution less skewed than with the scores collected in the interview. Therefore the calculation of a reliability coefficient was more feasible for the written list than for the other. From the total group of 792 cases, 100 cases were selected and the odd-even reliability was computed on the basis of the number of likes checked on the even numbered items and the number of likes checked on the odd numbered items. The corrected coefficient was .88. Very few psychological tests with only 22 items have reliabilities high enough to warrant their use with individuals.

Vocational interests are known to be influenced by age [3] and common sense suggests recreational interests are equally as, if not more, susceptible to the influence of increasing age. The ages of 780 of the 792 subjects were known. The mean age of this group was 24.4 years, with a standard deviation of 7.9 and a range from 17 years to 49 years. The total distribution was extremely skewed with 318 of the cases falling within the 17–18 year old group. The product-moment correlation of −.13 between age and range of interests does not offer a true estimate of this relationship because of the skewness.

[3] Strong, E. K., Jr. *Change of interests with age.* Stanford, Calif.: Stanford Univ. Press, 1931.

A better estimate of the influence of age upon range of interests was obtained by comparing the number of items checked as liked by the 406 men in the 17 to 22 year old group to the number liked by the 177 men in 29–49 year old group. The distributions for these two groups are presented in Table 2.

Table 2

Distributions of Number of Likes Obtained with Printed List for Different Age and Educational Groups

No. of Items Liked	17–22 Yr. Olds		29–49 Yr. Olds		17–22 Yr. Olds 8th Grade or Less		17–22 Yr. Olds H.S. Grad. or More	
	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
22	32	100	10	100	1	100	10	100
21	21	94	5	94	4	99	5	90
20	30	88	7	92	7	94	7	84
19	22	81	7	88	4	86	9	77
18	41	75	9	84	7	81	11	68
17	28	65	6	79	4	73	8	57
16	31	58	4	75	8	68	9	48
15	30	50	6	71	8	59	4	39
14	29	43	14	70	7	50	11	35
13	22	35	8	62	8	41	4	24
12	24	30	10	57	5	32	5	20
11	20	24	15	51	4	26	3	14
10	13	19	13	43	3	21	2	11
9	17	16	9	36	3	18	2	9
8	17	12	10	31	1	14	4	7
7	15	7	9	25	4	12	1	3
6	5	4	7	20	4	8	1	2
5	5	2	7	16	3	4		
4	3	1	12	12			1	1
3	1	1	5	5				
2								
1			4	2				
Total	406		177		85		97	

In the younger group, 50 per cent of the men liked 15 or more of the activities. In the older group, only 30 per cent liked 15 or more items. In the younger group, 12 per cent liked only 8 or fewer items, while in the older group, 31 per cent liked no more than this many activities. Inspection of the table reveals a substantial decrease in the number of activities liked as one moves from the younger to the older group. Table 3 presents

the per cent of people in each of these two age groups and in the groups divided upon the basis of education who check each item. In almost every case, the difference agrees with expectation.

Age brings a decreasing interest in dates. The younger men like to play horseshoes more than the older but this difference is influenced by a greater preference of the younger, less well educated group. Age apparently bears little influence on interest in dancing or on interest in playing cards. Younger men are more interested in parties, basketball, football and hiking. Older men are slightly more interested in watching baseball and in reading. Inspection of the table indicates that changes in interest accompanying age are not independent of the educational status of the individual.

The above results show that when an individual's range of interests is to be used as a personality index, his age must be considered. A man of forty years cannot be compared legitimately with a youth of eighteen.

Table 3

Percentage of Various Groups Checking that They Liked Each Item

Item	17–22 Yrs. 8th Gr. or Less	17–22 Yrs. More than 8th Gr.	29–49 Yrs. 8th Gr. or Less	29–49 Yrs. More than 8th Gr.	17–49 Yrs. 8th Gr. or Less	17–49 Yrs. More than 8th Gr.
Play checkers	71	62	50	62	61	63
Play pool	67	70	55	56	63	66
Play horseshoes	72	57	57	59	67	61
Dance	40	58	41	59	41	59
Go on dates	83	83	27	39	53	68
Box	46	48	22	32	36	42
Play cards	58	60	59	60	59	61
Go to parties	62	72	37	44	49	62
Play basketball	55	70	18	44	34	61
Play baseball	59	77	55	64	63	73
Play football	55	77	22	44	39	66
Fishing	76	76	71	78	76	77
Hiking	57	59	40	48	48	55
Go to carnivals	58	68	44	41	53	58
Go on picnics	54	62	52	49	53	58
Reading	63	71	63	82	63	75
Go to movies	90	95	77	88	86	94
Listen to radio	83	84	75	86	81	86
Roller skate	48	55	38	33	41	48
Watch baseball	70	71	75	79	75	73
Watch football	70	81	54	72	63	80
Go swimming	79	84	69	77	75	83
No. of cases	85	289	89	115	229	483

Table 3—*Continued*

Item	17–22 Yrs.	29–49 Yrs.	Inaptitude Discharge Cases	Normal Matched Controls
Play checkers	64	57	46	49
Play pool	69	55	24	62
Play horseshoes	61	58	32	67
Dance	54	51	27	47
Go on dates	83	34	27	50
Box	47	28	11	31
Play cards	59	60	37	61
Go to parties	69	41	24	48
Play basketball	67	32	14	37
Play baseball	73	60	36	65
Play football	71	34	26	39
Fishing	76	75	62	78
Hiking	59	45	17	37
Go to carnivals	65	42	27	44
Go on picnics	60	50	38	55
Reading	69	74	44	71
Go to movies	94	83	59	89
Listen to radio	84	81	79	81
Roller skate	53	35	16	46
Watch baseball	70	77	46	72
Watch football	78	64	39	66
Go swimming	82	74	34	83
No. of cases	374*	204*	144	144

* These total numbers of cases do not correspond with the totals presented in Table 2 as some cases for whom complete data were not available were discarded and additional cases were obtained to increase the size of the older group.

Separate normative tables must be used in order to provide a relevant frame of reference.

Range of interests might also vary with the educational status of the individual. In order to test this hypothesis a group of high educational attainment was compared with a group of lower educational attainment, holding age relatively constant. In the group of recruits from 17 through 22 years of age, complete data were avaiable for 85 men who had finished no more than the 8th grade and for 97 men who had at least graduated from high school.

In the higher education group, 52 per cent of the men liked 16 or more of the activities. In the lower education group, only 32 per cent liked 16 or more items. In the higher education group, 86 per cent liked 11 or more items, while in the lower education group, only 74 per cent liked

as many items. Table 2 presents the distributions for the two educational groups. Comparison of these differences with those found between the two different age groups reveals that much more variation in range of interests is related to changes in age than to differences in education, although the influence of education is nevertheless marked.

The total group for whom complete data were available was divided into those recruits who had completed the 8th grade or less and those who had completed more grades than this. Table 3 presents the per cent of 17-22 year old men and 29–49 year old men in each of these two educational groups who liked each of the listed items. The figures are also presented for all of the men who had completed no more than the 8th grade and all who had completed more than this.

Comparison of the two columns of figures for the two educational groups shows fewer changes in interests related to education than to age. The more educated men show more interest in dancing, in going on dates, in going to parties and in athletics associated with secondary schools. The less well educated groups like to play horseshoes more than the others. The differences in interests accompanying change in educational status well conform to the opportunities offered by our educational institutions.

In order to determine the relative influence of age and education upon responses to the items, the 22 items were ranked in order of popularity among the 17–22 year old men and among the 29–49 year old men. The correlation between these two rank orders was +.52, indicating substantial shifting of item popularity as one moved from the younger to the older group. The items were then ranked in order of popularity among the men completing 8 grades or less and among the men completing more than 8 grades. The correlation between these two rank orders was .78. These two correlations make plain that, although educational status affects item response, age differences are far more effective.

The author's impression is that age differentials must be allowed for in evaluating range of interests. Educational differences, however, need not be considered if convenience so dictates, although more exact information will be obtained if these differences are also taken into account.

The primary question here, as it was with the oral list, is how well does this list differentiate between people who are making a good adjustment to military training and people who are not.

Data were available for the three groups of non-adjusting individuals mentioned previously. The first group consisted of the 25 recruits selected by psychiatrists as requiring further study not involving intellectual level or literacy. The mean age of this group was 23.1 years with a standard deviation of 6.24 years. The second group consisted of the 23 recruits selected by psychiatrists as requiring examination concerning

possible mental deficiency or illiteracy. The mean age of this group was 26.6 years with a standard deviation of 5.76 years. The third group consisted of those 114 cases who were given inaptitude discharges. The mean age of this group was 25.7 years and the mean school grade completed was 8.8.

Distributions of the number of items liked by these different groups are presented in Table 4. Two different groups of normal recruits are presented for comparative purposes.

Table 4

Distributions of Number of Likes Obtained with Printed List for Various Groups of Normal and Non-Adjusting Recruits

No. of Items Liked	Screened for other than M.A. or Reading		Screened for M.A. or Reading		Aptitude Discharges		Normal Recruits not Screened		Normals Matched with Aptitude Discharges	
	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %	Freq.	Cum. %
22			2	100	1	100	46	100	6	100
21					2	98	34	94	3	94
20	1	100			2	97	45	89	4	92
19	1	96	1	92	1	95	39	83	6	88
18	2	92	1	88	2	94	64	78	11	83
17	2	84	1	82	1	93	38	69	1	74
16	1	76			2	92	45	64	9	73
15			2	78	2	90	42	58	4	65
14			3	70	4	88	54	52	7	61
13	2	72	1	56	3	85	38	45	4	55
12	1	64	1	52	2	82	51	39	8	52
11	4	60	1	48	2	81	47	32	11	45
10	1	44	1	44	4	80	39	26	5	35
9	2	40	4	39	7	75	29	21	11	29
8	2	32	2	22	11	70	28	17	3	21
7	2	24			11	60	32	13	7	18
6			1	13	10	50	16	8	5	12
5	1	16	1	8	11	41	17	6	3	8
4	2	12			14	31	17	4	3	5
3	1	4	1	4	12	19	7	2	3	3
2					10	8				
1							4	1		
Total	25		23		114		732		114	

The first group of 732 normal recruits includes all recruits for whom ages were available and who were not selected by psychiatrists as needing further examination. The mean age of this group was 25.4 years with a standard deviation of 4.0 years. The second group of normal controls consisted of 114 recruits selected from the larger control group and matched with the 114 discharged cases on the basis of age, highest grade completed, and age at leaving school. The effectiveness of this matching is revealed by comparing the mean age of 25.9 of this normal group and their mean grade attainment of 8.8 with the similar figures just presented for the 114 discharged recruits.

Comparisons of the distributions indicate that, with education and age constant, the non-adjusting individuals and the normals are differentiated by their ranges of interests as measured by this list. When a group of individuals is tested and then from this group, people are selected by psychiatrists as requiring further examination, this selected group will have restricted ranges of interests. Similarly, when a group, already studied by psychiatrists and psychologists and found to have an unfavorable prognosis, is tested, people in this group will also have restricted ranges of interests.

The mean scores of the two matched groups were compared. The mean score of the 114 normal cases was 12.8 with a standard deviation of 5.2. The mean score of the 114 discharged cases was 6.9 with a standard deviation of 5.0. The critical ratio of the difference between these two means, based on the standard error of the difference and taking into account the low but positive correlation between the groups, was 9.7. Beyond all doubt the scale differentiates between the two groups.

In the discharged group, 50 per cent liked six or fewer items,while in the normal group, only 12 per cent liked 6 or fewer items. In the normal group, 35 per cent liked more than 15 items while in the discharged group only 10 per cent like as many items. The relatively small amount of overlapping indicates the list measures a trait which clearly differentiates those people who have adjusted to military training and those who have not and this trait can be measured with sufficient exactitude to be useful in individual cases.

Conclusion

The preliminary work with the orally presented list and the later work with the printed list show that the range of interests offers an indirect and convenient method for predicting adjustment to military training. The results show more overlapping between groups with the printed list than with the oral list. The more careful control of age and educational factors obtained in the analysis of the printed list, however, must be

considered. This and the greater convenience and objectivity of the printed list give it added advantage.

The method can be useful in several situations. First it may be used as a screening device. As it does not select mental defectives and perhaps many other groups aimed at in psychiatric screening, the list is in itself not a sufficient screening technique. Hunt, Wittson and Harris [4] have concluded that screen tests are most useful to the military psychiatrist in supplementing the results of his clinical interview. Most acceptable is their suggestion that screening tests, such as developed here, and the psychiatric interview are best used to complement each other.

The method is also useful in the collection of information to be used in making the final judgment as to whether a recruit should be retained in military service or not. The validity of these judgments is dependent upon the reliability, the relevancy and the amount of information upon which they are based.

The analysis leaves little question that groups not adjusting well to military training tend to have significantly restricted ranges of interests and this restriction can be measured reliably in individual cases and used in making prognostic judgments.

Received July 17, 1944.

[4] Hunt, W. A., Wittson, C. L., and Harris, H. I. The screen test in military selection. *Psychol. Rev.*, 1944, 51, 37–46.

A Note on the Problem of Brain Damage in Rehabilitation and Personnel Work

Howard F. Hunt

University of Minnesota

Brain injury in the adult human being is usually followed by alterations in personality and intellectual ability as well as by alterations in physical status. The symptomatology and vocational implication of these alterations deserves discussion since many returning brain-injured veterans will be only partially incapacitated and will desire some means of self-support. The question of the employability of brain-injured persons is an important one at the present time, moreover, since this condition is not too uncommon among the general population as a result of accidents, disease, and excessive use of alcohol. This discussion is intended to provide personnel and rehabilitation workers who have not specialized in clinical psychology with useful information for dealing with brain damage cases. Only the psychologic symptoms of greatest practical import from the standpoint of vocational guidance and personnel work will be discussed here. No attempt will be made to discuss the "basic, underlying defect, or defects" of which these more or less superficial symptoms are a function.

The residuals of brain injury may, for purposes of convenience, be classified under two headings: primarily physical or somatic symptoms and symptoms which are primarily behavioral or psychologic. Among the more important physical symptoms are included: weakness and fatigability, convulsive disorders, sensory and motor disorders (including the aphasias), and defects in muscular coordination. These symptoms are generally detectable by routine physical and neurological examinations and from the medical history obtained from such a person. With the possible exception of the aphasias, the somatic disabilities usually come within the purview of the physician and will not be discussed here.

The psychological symptoms consist superficially of intellectual deterioration and alterations in the individual's emotional life and personality. The severity of the symptoms depends largely upon the severity and extent of the damage incurred and upon the prior personality organization and intellectual level of the affected individual, varying quantitatively from exceedingly subtle alterations in persons who superficially seem quite normal to gross dementia with disorientation as to

time, place and even person. In some cases both the psychologic and somatic symptoms may be present in varied degrees and combinations, while in other cases either the psychologic or the somatic symptoms alone may be appreciable. Even relatively small areas of damage in certain locations may produce distinct physical symptoms but little or almost no psychologic alteration, whereas lesions placed in other areas of the brain may result in behavioral defects unaccompanied by appreciable somatic abnormality. We are primarily concerned with the latter type of case in this discussion.

The intellectual deterioration or loss of intellectual ability resulting from brain damage apparently involves those more dynamic aspects of intelligence which are crucial for problem solving. Brain-damaged persons commonly show some loss of memory for recent events, a symptom which is related to their decreased ability to learn unfamiliar material or new tasks. They also show rigidity and perseveration in their mode of attack on new problems and in their behavior in general. This defect is manifested in their decreased ability to organize or synthesize new behavior patterns and in their tendency to continue to apply old, familiar methods of solution to new problems even after the method has proved unsuccessful. This characteristic may appear in social situations in the guise of decreased ability to follow rapid shifts in the form and content of conversations and diminished skill in coping with sudden turns of events. In addition to these symptoms, a general slowing of thinking and speed of reaction as well as an increased distractibility are often observed. In general the old, familiar response patterns, habits, and skills are relatively unaffected by mild to moderate degrees of brain damage.

Alterations in the individual's emotionality and, consequently, in personality frequently follow brain damage. Emotional reactions tend to be more easily aroused and to greater heights in the brain-damaged person than in the normal person, and they tend to abate more rapidly once aroused. Thus, these people often give the impression of a relative lack of inhibition or emotional control as well as of deficient foresight and concern about their own future or the consequences of their acts. Many show a mild degree of perhaps unjustified optimism and buoyancy or euphoria, while some others may show a tendency toward an exaggeration of their previous personality characteristics. These symptoms, plus the accompanying intellectual changes, set the stage for childish or antisocial and unethical behavior in predisposed persons. In general, the individual's power of adjustment to society and to the problems of everyday life is diminished.

No specific personality "type" seems to be invariably peculiar to or characteristic of brain-damaged persons. In the case of adults, the de-

fects described above are imposed on a matured personality "system" or structure. The resultant behavior characteristics will thus be a product of the interaction between the individual's original personality and intellectual level and the specific residuals associated with brain damage, plus the frequently exaggerated reactions of this modified personality to the stresses and strains incidental to the demands of daily life. This latter factor may be particularly important in some cases since the brain-injured person's goals and expectations may remain relatively unaltered though his capacity for achieving them is reduced. Thus, a person whose powers of adaptation and emotional control are diminished must often live under and adjust to particularly trying circumstances. It is not surprising that some brain-injured persons develop psychoneuroses or exhibit socially unacceptable behavior. Sensori-motor and speech defects or convulsive disorders may also complicate this situation, as may psychoneurotic reactions on the part of the affected person to the idea of being handicapped or to the circumstances associated with a spectacular, injury-producing accident or illness.

Because there is no regeneration of destroyed neuronal tissue within the brain, the permanence of the defects associated with brain damage has sometimes been assumed. Partial and sometimes almost complete clinical recovery does occur in some cases, however; but in which cases it will occur cannot yet be predicted accurately. Recovery, if it is to occur to an appreciable degree, is usually well under way within a period of a few months after injury and is not generally to be expected in cases of long standing deterioration.

This discussion has been primarily concerned with the symptoms of cases in which brain damage occurs after the individual has attained almost full intellectual growth, that is after the person is fifteen or sixteen years old. The defects resulting from damage earlier in life are somewhat similar with additional complications arising from arrested or uneven intellectual and personality development; consequently defective education, often with marked and specific deficiencies in one or more of the tool subjects such as arithmetic or reading; and the development of various compensatory habits and unacceptable reaction patterns. Probably only the less severely affected cases of this type will be encountered in the employment situation since, in most cases, the segregating and selecting influences of school and society have eliminated the grossly unfit from the active labor market.

The vocational prognosis attached to brain-damaged persons depends upon a number of factors. The severity of the handicap in such persons is relative to their original intellectual level and personality. It depends not only upon the extent and placement of the damage and upon the

absolute magnitude of the intellectual and emotional alterations but also upon the individual's original intellectual ability, education and training, and the excellence of his personality integration and adjustment. The magnitude of the handicap is also relative to the type of work he wishes to perform and to his interest in and familiarity with it.

The exact vocational implication of mild to moderate psychologic defect following brain damage is as yet unknown, but a few generalizations may be drawn from our relatively limited clinical experience with it and on the basis of the symptoms of the condition. Occupations placing a premium on physical stamina, inventiveness and ingenuity, emotional control, and diplomacy in inter-personal relations are probably contra-indicated in the majority of cases just as are positions involving great personal responsibility both for the welfare of other persons and for vital programs or equipment. Special experience and training or ability and good personal adjustment coupled with relatively minimal intellectual and emotional changes, however, would warrant exception to the above generalizations in some cases. Usually, nevertheless, even mild psychologic defect is probably an unfavorable prognostic sign with regard to both vocational and personal adjustment.

Though diagnosis and treatment of brain damage are primarily medical problems, the standard techniques utilized by physicians fail to yield certain types of information crucial to the vocational management of affected persons. Medical facilities for detecting and estimating the severity of the psychologic symptoms associated with brain damage consist largely of unrefined, uncalibrated tests aided by subjective, clinical impressions and estimates. These techniques yield relatively unreliable estimates of degree of deterioration. Also, since they are relatively insensitive, they may fail to detect minimal defects, especially in cases where the physical symptoms are not appreciable. In a few cases, moreover, the physical and neurologic examination findings may be entirely negative, the psychologic defects minimal, and the brain damage detectable only by refined psychometric methods. Recent advances in clinical psychology have provided more adequate and more sensitive methods. Though these methods do not "measure" in the strictest mathematical sense, they represent a marked improvement over the traditional medical techniques in that they are somewhat more sensitive and do yield rough estimates of degree of deterioration. They are particularly useful in evaluating the employable brain-damage cases—those with relatively minimal defect.

Clinical psychology has, however, no special technique to offer for the detection of minimal alterations in the emotionality of brain-damaged persons. The standard personality tests are of considerable value in

assessing the actual personality characteristics of these people, nevertheless. A personality deviation in a brain-damaged person may have more serious implications than a similar deviation in a person who is intact.

To be of practical value in the personnel or vocational guidance situation, a psychometric instrument must not require extended administration time or extensive, cumbersome equipment, and yet it must be maximally sensitive. Ease of administration and scoring as well as interpretability are also crucial. Of course, the test must have clinical validity and be applicable to the type of person being tested, applicable both from the standpoint of its content and from the standpoint of its standardization. As is well known, however, the use of tests provides no royal road to diagnosis. The pitfalls for the unwary are many, and the thoughtless clinician all too easily may be lulled into a false sense of security by the arithmetic form and so-called "objectivity" of test results. A test is no better than the clinician who uses it, and an otherwise useful test may be rendered ineffective by thoughtless, careless administration and naive interpretation. The examiner must be thoroughly familiar with the administration technique as well as fully cognizant of the psychologic aspects of the task and total situation with which the tested person is confronted. This is particularly important in brain-damaged cases because of their characteristic emotionality and psychologic defects. Naturally, test results must be considered as but a part of the evidence required for diagnosis and must always be interpreted in the light of all the rest of the data relevant to the case being studied.

The standard intelligence tests have often been used to detect intellectual deterioration, but they are somewhat ineffective tools for this purpose from the strict psychometric standpoint because of their relative insensitivity to this condition. They may, however, provide the occasion for behavior of great diagnostic significance to an experienced clinical psychologist. In general, the same evaluation applies to the Rorschach (ink blot) method when used as a deterioration test. Moreover, much specialized experience and training are required for proper administration, scoring, and interpretation of this procedure. Accordingly, the use of psychometric instruments specially designed for the detection of deterioration is advised, particularly in borderline cases.

A number of deterioration tests have been developed, but the more effective of such tests have been based upon what might be called the differential score technique which was originally suggested by Babcock (1). This technique involves measuring the discrepancy between a person's performance on tests sensitive to brain damage (tests of learning, memory, abstract thinking) and his performance on tests relatively insensitive to this condition (tests of vocabulary, information). Significant discrepan-

cies indicating impaired performance on the sensitive tests relative to performance on the insensitive tests would thus suggest brain damage. This neat situation is complicated, however, by the tendency of depressed, anxious, schizophrenic, and other varieties of disturbed persons to obtain similarly discrepant scores on the two types of tests. Thus, such pathologic discrepancies, or deterioration scores, are ambiguous diagnostically and should be confirmed by case history and neurologic examination data before anything stronger than a suspicion of brain damage is justified.

In cases with pathologic deterioration test scores, differentiation as to the etiology of the score is important from the standpoint of long-time vocational prognosis. Pathologic scores attributable to emotional-motivational disturbances as in the depressions and the anxiety states indicate an impairment of intellectual efficiency which usually disappears when the afflicted individuals recover. Schizophrenic deterioration generally disappears in a like manner in those cases which do recover. In contrast, the intellectual impairment associated with brain damage is usually considerably less transient and is accompanied, therefore, by a less favorable vocational prognosis. Because of the proneness of brain-damaged persons to emotional disturbances, the more permanent impairment associated with the brain damage may be temporarily augmented by depression, anxiety, and the like. Under these conditions, the ultimate intellectual status of the person is difficult to determine since some of the impairment indicated by the test score may be expected, in many cases, to disappear when emotional equilibrium is regained.

Deterioration test scores are thus not a final index of vocational prognosis but rather a diagnostic and prognostic aid. The extent to which they aid diagnosis and prognosis depends to a substantial degree upon the skill and clinical acuity of the interpreting clinician.

Of the available, specially devised deterioration tests for adults, the Shipley-Hartford Deterioration Test (8) is the easiest to administer. It is a twenty minute, self-administering, paper and pencil test which may be used only with cooperative, test-sophisticated subjects of average or above average intelligence.

The Babcock-Levy Revised Test for Efficiency of Mental Functioning (2) is an individual test very similar to the standard individual intelligence tests in the time, skill, and amount of equipment required for administration. It is applicable to cooperative persons of a somewhat less restricted range of intelligence and test-sophistication than is the Shipley. The norms for these two tests include no correction for normal deterioration with advanced age, so subjective correction for this factor must be made in the interpretation of the test performances of older persons.

The Hunt-Minnesota Test for Organic Brain Damage (4) is an indi-

vidual test requiring some practice but relatively little equipment for administration. The test can be completed in from fifteen to thirty five minutes, depending upon whether the short or the long form is used. It is applicable to cooperative adults with mental ages of over eight years, and the norms include a statistical correction for normal age changes within the range of sixteen to seventy years. Scores on this test obtained from persons near the extremes of the age and intelligence ranges must be interpreted with considerable caution, however.

"Psychological deficit" has been suggested (6) as an inclusive but neutral term for the impairment of intellectual efficiency associated both with brain damage and with emotional-motivational disorders. The three tests described above are apparently fairly effective in detecting this deficit. In the development of the Hunt test, an attempt was made to provide a special means for identifying those pathologic scores attributable to emotional-motivational disturbances so that the test would then be a specific test for the deterioration associated with brain damage. This attempt has been only partially successful. Since neither the Shipley nor the Babcock tests make this discrimination, pathologic scores on any of these three tests may indicate either brain damage or an emotional-motivational disturbance accompanied by poor cooperation, motivation, and attention. Consequently, great care must be exercised in the interpretation of scores obtained on these tests.

Received July 18, 1944.

References

1. Babcock, H. An experiment in the measurement of intellectual deterioration. *Arch. Psychol.*, N. Y., 1930, No. 117.
2. Babcock, H., and Levy, L. *Test and manual of directions: the revised examination for the measurement of the efficiency of mental functioning.* Chicago: C. H. Stoelting Co., 1940.
3. Hunt, Howard F. A practical, clinical test for organic brain damage. *J. appl. Psychol.*, 1943, 27, 375–386.
4. Hunt, Howard F. *The Hunt-Minnesota test for organic brain damage.* Minneapolis: The University of Minnesota Press, 1943.
5. Hunt, Howard F. A note on the clinical use of the Hunt-Minnesota test for organic brain damage. *J. appl. Psychol.*, 1944, 28, 175–178.
6. Hunt, J. McV. *Personality and the behavior disorders.* New York: The Ronald Press, 1944.
7. Pollack, B. The validity of the Shipley-Hartford Retreat test for "deterioration." *Psychiat. Quart.*, 1942, 16, 119–131.
8. Shipley, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, 9, 371–377.
9. Shipley, W. C., and Burlingame, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.*, 1941, 97, 1313–1324.

Personality Patterns of Adolescent Girls: II. Delinquents and Non-Delinquents

Dora F. Capwell

Trainee Acceptance Center, Public Schools, Pittsburgh, Pa.

In a study devoted to the personality patterns of adolescent girls who show improvement in IQ (2) two groups of girls were used as subjects, a group of delinquents and a group of non-delinquents. A series of personality tests were administered to each group on two occasions along with tests of intelligence and academic achievement. The results of the personality tests, aside from their contribution to the original problem, are of special interest due to the degree and manner in which they differentiated the delinquents from the non-delinquents. Most present day students of delinquency agree with the point of view expressed by Lowrey (10). "Delinquency is probably most frequently due to the subtle effects of interactions between individual and environment, leading to the establishment of particular personality sets." However, there has been disagreement regarding the success with which one can measure and delineate these personality sets by means of standard, objective tests of personality. The following results show the differences of personality between a group of delinquent girls and a group of non-delinquent girls as measured by one series of personality tests.

The procedure has been described in detail in a previous report (2). A total of 101 delinquent girls at the Minnesota State School for Girls and 85 non-delinquents in the Public Schools of Sauk Centre, Minn., were given a psychological examination and re-examined from 4 to 15 months later. The personality tests which were given twice were the Minnesota Multiphasic Personality Inventory (6), the Washburne Social Adjustment Inventory (13) and the Pressey Interest-Attitude Test (11). Two other tests of personality, the Terman-Miles Test of Masculinity-Feminity (12) and the Vineland Social Maturity Scale (3), were given just once. The levels of intelligence and academic achievement were determined by the Kuhlmann Tests of Mental Development (9) and the Stanford Achievement Test (8).

Results

Before examining the personality test results, it is important to note the differences between the two groups in intelligence and achievement.

The non-delinquents were girls of higher intelligence, showing a mean IQ on the first test of 101, as compared with a mean of 87 for the delinquents. The standard deviation for each group was 17. Each group showed some retardation in school achievement when compared with the norms of the Stanford Achievement Test. The delinquents were more retarded than the non-delinquents, as would be expected, but the differences are not statistically reliable. The amount of retardation in terms of grade scores is shown in Table 1.

Table 1

Difference between Actual Grade Level and Achievement Grade Score *

Score	Delinquents $N = 97$**		Non-Delinquents $N = 85$		$D/\sigma D$ between Delinquents and Non-Delinquents
	Mean Diff.	S.D.	Mean Diff.	S.D.	
Total Score	−1.30	1.41	−.82	1.50	2.28
Reading Score	−.55	1.52	−.49	1.46	.28
Arithmetic Score	−1.96	1.76	−1.28	2.03	2.42

* 1.00 equals one total grade.

** Four delinquents were absent when these tests were given.

The personality tests discriminated the delinquents from the non-delinquents with varying success; two of them showed striking differences. The significance of differences between the two groups on the Minnesota Multiphasic Personality Inventory is shown in Table 2.

Table 2

Minnesota Multiphasic Personality Inventory—Significance of Difference of Raw Scores

Scale	Difference between Delinquents and Non-Delinquents	
	First Test $D/\sigma D$	Second Test $D/\sigma D$
"?"	2.93	4.25
L	1.75	3.30
F	7.21	5.95
Hs	3.11	3.12
D	4.59	2.92
Hy	2.74	.60
Pd	16.00	14.00
Pa	12.00	8.03
Pt	6.64	7.36
Sc	7.10	8.55
Ma	8.00	7.95

Each scale except the Hy, or Hysteria scale, shows a clear differentiation between the two groups. The greatest difference appears in the scores for Pd (Psychopathic Deviate). The extent to which the Multiphasic differentiated the groups may be seen in Table 3, and Figures 1 and 2 show graphically the differences in average T-scores between the two groups.

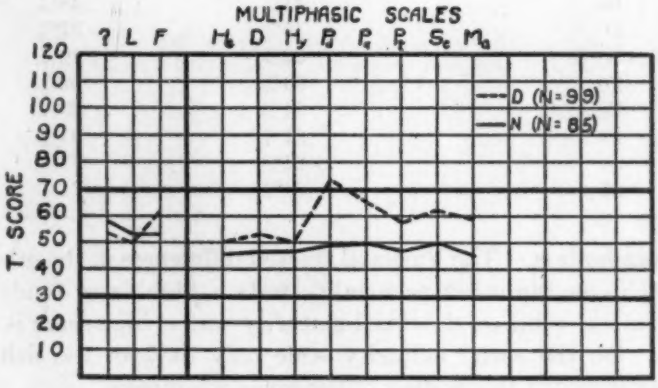

FIG. 1. T-score profile on the Multiphasic Inventory for the Delinquent (D) and Non-Delinquent (N) groups, 1st test.

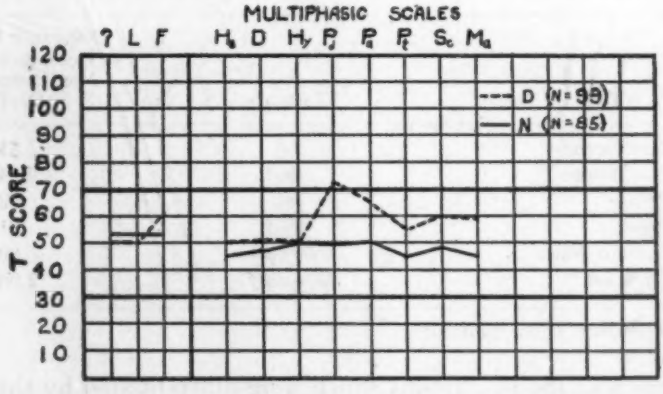

FIG. 2. T-score profile on the Multiphasic Inventory for the Delinquent (D) and Non-Delinquent (N) groups, 2nd test.

Table 4 shows the significance of differences between the raw scores of the other personality tests. The Washburne and the Vineland differentiated between the delinquents and non-delinquents, but the Pressey Interest-Attitude and the Terman-Miles did not. On the Washburne 54% of the delinquents reached or exceeded the 75th percentile of the non-delinquents' scores on the first examination, and 51% did so on the

Table 3

Percentage of Overlap on the Multiphasic

Multiphasic Scale	Percentage of Delinquents who Reached or Exceeded the 75th Percentile of Non-Delinquents	
	1st Test	2nd Test
Hs	41%	48%
D	48%	37%
Hy	32%	19%
Pd	93%	92%
Pa	84%	65%
Pt	61%	61%
Sc	64%	71%
Ma	64%	71%

second examination. The Vineland showed differences in the other direction, in that on the other personality tests a high score tends toward maladjustment, while on the social maturity scale a high score is a favorable one. On the social maturity scale only 31% of the delinquents reached or exceeded the median of the non-delinquents.

Table 4

Other Personality Tests—Significance of Difference of Raw Scores

Test	Test No.	Difference between Delinquents and Non-Delinquents $D/\sigma D$
Washburne	1st	5.31
Washburne	2nd	4.54
Pressey	1st	.62
Pressey	2nd	.61
Terman-Miles	(one only)	.89
Vineland *	(one only)	2.96

* $D/\sigma D$ between social quotients.

Inasmuch as the two groups which were differentiated by the Multiphasic, the Washburne, and the Vineland differ in level of intelligence as well as in regard to delinquency, it is necessary to find out if the personality test differences are related to the differences in intelligence. Hence, 52 delinquents from the group were matched within two IQ points with 52 of the non-delinquents in order to compare personality differences and percentage of overlap on those tests which differentiated the groups reliably. The mean IQ of each group was 95 with a standard deviation of 14. Only those tests were compared which differentiated the larger

groups. Results obtained with the matched groups may be seen in Table 5.

The Multiphasic continued to differentiate the delinquents and non-delinquents reliably with the exception of the Hs scale. Hy was not used because it did not differentiate the larger groups. The Washburne fell

Table 5

Matched Groups of Delinquents and Non-Delinquents: Significance of Difference of Raw Scores on First Examination and Percentage of Overlap

Test	D/σD	Percentage of Delinquents who Reached or Exceeded 75th Percentile of Non-Delinquents
Multiphasic		
Hs	1.95	—
D	3.91	55%
Pd	10.34	93%
Pa	8.43	84%
Pt	4.86	57%
Sc	4.06	57%
Ma	5.22	63%
Washburne	2.70	50%
Vineland	.38	—

slightly below satisfactory reliability of differences, and the Vineland failed to demonstrate any real difference when the groups were matched for IQ. Hence, the differences on the Vineland appear to be related more to intelligence than to delinquency, but the differences on the Washburne and the Multiphasic continue to be significant.

Discussion

The test which differentiated most clearly between these two groups was the Multiphasic. The results, which include the three validating scores, show first of all that even quite young adolescents do answer a well-devised personality inventory validly. The mean validity scores are close to a T-score of 50 for all groups, except for the delinquent scores on the F-scale; these were higher, but not sufficiently so as to invalidate the results. They would be expected to be somewhat higher than the F-scores for the non-delinquents because of the much greater amount of maladjustment in the delinquent group, and since, as stated before, there is a slight tendency for the F-score to be higher as gross maladjustment increases. The non-delinquents had many more "?" responses than the delinquents, due probably to two factors: first, they were a slightly younger group and found more items they did not understand, and secondly, the delinquents

had more leisure at the time of taking the tests. On the whole, however, the results from both groups showed a reassuring degree of validity.

It has long been believed that delinquents are more generally unstable than the normal population, and the Multiphasic Inventory bore out this belief by showing a significant difference of mean scores between delinquents and non-delinquents on seven of the eight scales, the delinquents scoring further away from the mean for the normal population. Even when the scores were not equivalent to a T-score of 70 or over, the criterion for significant maladjustment, they still were further toward the maladjustment end of the scale than were the scores of the non-delinquents. The most conspicuous differences were on the scales of Psychopathic Deviate and Paranoia. The delinquents had a mean T-score of 73 on the Pd scale and 65 on the Pa scale.

The delinquents were not as depressed as one would imagine, on their arrival at a correctional institution. That they are not is a finding which is consistent with the theory of the psychopathic deviate set forth by Hathaway (6), suggesting that the psychopathic inferior or deviate has trouble in his relations to society partly because he does not react with the emotions with which the normal person reacts under similar circumstances. Therefore, although the newly committed delinquent may seem to feel very badly, cry noisily, and appear to be extremely depressed, the test scores and also the speed with which she often recovers from this mood attest to the belief that it is a relatively superficial emotion. In the non-delinquent group only two girls made scores on the Pd scale as high as the mean for the delinquent group. Of these two one was a girl so maladjusted that she was being considered for foster-home placement by the child welfare worker of the county, and the other was a girl who had left school for a time, returned of her own volition, but was again considering leaving and was so upset that she was of concern to the school principal; she herself asked for a conference with the examiner to talk about her difficulties. Thus, this test measures adjustment in the same sense in which non-delinquents are better adjusted than delinquents.

Two other personality tests, the Pressey Interest-Attitude Test and the Terman-Miles Test of Masculinity-Femininity, when their raw scores were compared as with the other tests, did not discriminate between the delinquents and non-delinquents. This is contrary to the results implied by the report of Durea and Fertman (5), who gave the Pressey to 180 delinquent girls. Their study merely showed, however, that the scores from the delinquents compare "unfavorably with norms for non-delinquents." No control group was used at the same time. In the present data the delinquents compare unfavorably with the *norms*. but so does this particular sample of normal cases. These same normal girls made

normally expected scores on the other tests. The Pressey apparently does not measure the type of adjustment measured by the Multiphasic or the Washburne. The Terman-Miles did not show significant differences between the groups, but each group contained some cases of extreme scores, as one may judge from the standard deviation of the groups. The deviate cases were not always the girls who tested as most maladjusted on other tests.

The Vineland Social Maturity Scale did not show as much difference in social maturity as might be expected between the delinquents and non-delinquents. These girls do not show as much social retardation as the delinquent boys reported by Doll and Fitch (4), but there are other marked differences between the boys they studied and the girls of this study, the most important one being that the boys were much more retarded mentally than the girls are. The median mental age for the boys was 9.3 years, as against a median life age of 14 years. The social quotients were nearer the mental ages than the life ages. In the present group of delinquent girls the mean social quotient is 4 points below the mean IQ and is more related to intelligence level than delinquency.

In concluding the discussion of these test results a word may be said about the use of personality tests as measures of adjustment both with groups and individuals. Whether or not one considers them valid and helpful instruments to aid diagnosis appears to depend largely on the selection of tests one uses. Boynton and Walsworth (1) used six tests, all different from the ones used in the present study, with 47 delinquent and 50 normal girls and found that only one score of one test yielded a C.R. of 3.00 between the groups. They concluded that the tests "in the main do not provide empirical evidence of sufficient validity to justify one in putting a great deal of faith in them in individual and group diagnosis." They concluded further that since there was such disparity between the test results, delinquent behavior is not necessarily associated with personality aberrations. Somewhat different conclusions are warranted from the results of tests used in the present study. Two of them, the Terman-Miles M.-F. Test and the Pressey Interest-Attitude Test, were not helpful in differentiating delinquents and non-delinquents. Two others, however, gave satisfactory differences and should be helpful in either individual or group diagnosis; these two were the Minnesota Multiphasic Personality Inventory and the Washburne Social Adjustment Inventory. The results of both suggest that personality aberrations frequently *are* associated with delinquency. The use of personality tests with any special group necessitates careful selection of tests on the basis of the way they were standardized as well as their suitability for providing the desired information. Beyond that, experience with the results in

the particular group concerned is necessary before a fair evaluation can be made of the usefulness of the test. Generalization from a selected group of personality tests to all tests of that type is not warranted any more than from one intelligence test to another.

Summary and Conclusions

A group of 101 delinquent and 85 non-delinquent girls were tested with a battery of personality tests and retested from 4 to 15 months later. Differentiation between the two groups was measured by computing the significance of difference of the mean scores for each group and also the percentage of overlap. The groups differed in level of intelligence as well as in delinquent tendencies, so the effect of mental level on the personality test scores was investigated by similar statistical treatment of the scores of 52 girls from each group who were matched for IQ. The results led to the following conclusions:

1. The Minnesota Multiphasic Personality Inventory and the Washburne S.-A. Inventory discriminated the delinquents from the non-delinquents in degree of personality adjustment.

2. The Vineland Social Maturity Scale showed differences which were more related to intelligence than to delinquency.

3. The Pressey Interest-Attitude Test and the Terman-Miles Test of Masculinity-Femininity did not discriminate the delinquents and non-delinquents.

4. Personality tests may be of value with individuals or groups in measuring and describing the personality patterns of delinquent as distinguished from non-delinquent girls.

Received June 12, 1944.

References

1. Boynton, P. L., and Walsworth, B. M. Emotionality test scores of delinquent and non-delinquent girls. *J. abnorm. soc. Psychol.*, 1943, 38, 87–92.
2. Capwell, Dora F. Personality patterns of adolescent girls: I. Girls who show improvement in I.Q. *J. appl. Psychol.*, 1945, 29, 212–228.
3. Doll, E. A. A genetic scale of social maturity. *Amer. J. Orthopsychiat.*, 1935, 5, 180–188.
4. Doll, E. A., and Fitch, K. A. Social competence of juvenile delinquents. *J. of crim. Law Criminol.*, 1939, 30, 52–67.
5. Durea, M. A., and Fertman, M. H. Emotional maturity of delinquent girls. *Amer. J. Orthopsychiat.*, 1941, 11, 335–338.
6. Hathaway, S. R. The personality inventory as an aid in the diagnosis of psychopathic inferiors. *J. consult. Psychol.*, 1939, 3, 112–117.
7. Hathaway, S. R., and McKinley, J. C. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *J. Psychol.*, 1940, 10, 249–254.

8. Kelley, T. L., Ruch, G. M., and Terman, L. M. *Stanford achievement Tests: Manual of directions.* Yonkers-on-Hudson, New York: World Book Company, 1940.
9. Kuhlmann, F. *Tests of mental development.* Minneapolis, Minn.: Educ. Test Bureau, 1939.
10. Lowrey, L. G. Delinquent and criminal personalities. In J. Mcv. Hunt (Ed.). *Personality and the behavior disorders.* New York: The Ronald Press Company, 1944, Vol. II, pp. 794–821.
11. Pressey, S. L., and Pressey, L. C. Development of the interest-attitude tests. *J. appl. Psychol.*, 1933, **17**, 1–16.
12. Terman, L. M., and Miles, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.
13. Washburne, J. N. A test of social adjustment. *J. appl. Psychol.*, 1938, 19, No. 2, 125–144.

The Construction of the Guilford-Martin Inventory of Factors G-A-M-I-N

Howard G. Martin

Los Angeles, California

Factor analysis methods have been applied to the problem of isolating independent variables of temperament with the result that several unitary traits have been identified (3, 4, 5) and a series of inventories constructed which attempt to measure some of these traits (1, 6).

The Guilford-Martin Inventory of Factors G–A–M–I–N adds five more temperament variables to the eight already covered by the two preceding tests of the series. The five traits included in the new inventory are: **G**, General pressure for overt activity; **A**, Ascendancy in social situations as opposed to submissiveness; leadership qualities; **M**, Masculinity of attitudes and interests as opposed to femininity; **I**, Lack of inferiority feelings; self-confidence; and **N**, Lack of nervous tenseness and irritability.

Traits G, M, and N were discovered in factor analyses by Guilford and Guilford (3, 5) and trait I by Mosier (7). Traits A, M, I, and N have been recognized by clinical psychologists, though not always defined in the same manner. The definitions of the trait names as used in this connection are operational; that is, in terms of factorial studies and subsequent item analyses.

More than 300 questionnaire items were constructed which were believed to cover the areas of behavior constituting these five traits. This list, stated in question form to be answered by either "Yes," "?," or "No," was administered to 250 men and 250 women college students between the ages of 19 and 30, ranging from sophomores to seniors in southern California colleges and universities.

The items which had been shown by the factor analyses and two previous item analyses to have heavy loadings in a trait were included in the preliminary scoring key for that trait. Typical items from these keys are listed below:

For trait G:

Are you inclined to be quick in your actions?
Can you turn out a large amount of work in a short time?
Are you inclined to rush from one activity to another without pausing for rest?

For trait A:
Do you usually speak out in a meeting to oppose someone who you feel sure is wrong?
Do you find it difficult to get rid of a salesman to whom you do not care to listen or give your time?
Have you ever, on your own initiative, organized a club or group of any kind?
For trait M
Do you like love scenes in a movie or a play?
Do you (or would you) like to go hunting with a rifle for wild game?
Are you disgusted at the sound of foul language?
For trait I:
Do you feel that the average person has made a better adjustment to life than you have?
Do you feel confident that you can cope with most situations that you will meet in the future?
Do you suffer keenly from feelings of inferiority?
For trait N:
Do you often become irritated over little annoyances?
Do you have nervous habits such as chewing your pencil or biting your fingernails?
Do you feel compelled to change your bodily posture frequently while sitting?

Four hundred of the questionnaires (200 males and 200 females) were scored with the preliminary scoring keys thus constructed. The highest 100 and the lowest 100 cases (extreme quarters of the distributions of scores) were used as criterion groups in the item analyses for factors G, A, I, and N. For the item analysis of factor M, the criterion groups consisted of the 100 male cases highest on the preliminary scoring key for the M factor and the 100 female cases lowest on the key.

Scoring weights were found for each response to each item by using Guilford's *abac* method (2). This procedure yielded final scoring keys consisting of 41 items for trait G, 50 items for trait A, 52 items for trait M, 64 items for trait I, and 68 items for trait N. Only nine items were scored for more than one trait.

Table 1

Intercorrelations between Scores of 100 Cases

	G	A	M	I
A	.51			
M	.16	.34		
I	.39	.54	.43	
N	−.27	.16	.34	.70

As a check on the reliability of scoring, the test papers of the remaining 100 cases, including 50 males and 50 females, were scored with these final scoring keys, which, for the purpose, were divided into random halves of items. Pearsonian coefficients of correlation were computed between

the scores on these halves of the scoring keys and, when corrected by the Spearman-Brown formula, they became .89 for trait G, .88 for trait A, .85 for trait M, .91 for trait I, and .89 for trait N.

Intercorrelations between the scores derived by means of these scoring keys are shown in Table 1.

The validity of trait M was checked by comparing the distributions on the trait scores of the 50 males and the 50 females which were not used in the item analyses. Ninety-two per cent (46 out of 50) of the males were above the median of the distribution of the 100 scores of the two sexes combined. Ninety-two per cent (46 out of 50) of the females were below the median of this distribution. The overlap was only 8 per cent of the 100 scores. Some males are more feminine in their reactions than some females and vice versa (8). That this difference in some cases may have a glandular condition accompanying it is demonstrated by the fact that one of the males who fell below the median on trait M was found to have a history of hypogonadal function and infantile genitalia.

These types constitute an undetermined percentage of the overlap in scores between the sexes in general. Therefore, the validity coefficient (phi) of .84 for trait M which was computed from these figures is as high as can be expected in view of the fallibility of the criterion.

Received July 22, 1944.

References

1. Guilford, J. P. *An inventory of factors S-T-D-C-R.* Beverly Hills, California: Sheridan Supply Company, 1940.
2. Guilford, J. P. A simple scoring weight for test items and its reliability. *Psychometrika*, 1941, **4**, 367–374.
3. Guilford, J. P., and Guilford, R. B. Personality factors S, E, and M, and their measurement. *J. Psychol.*, 1936, **2**, 107–127.
4. Guilford, J. P., and Guilford, R. B. Personality factors D, R, T, and A. *J. Abn. & Soc. Psychol.*, 1939, **34**, 21–36.
5. Guilford, J. P., and Guilford, R. B. Personality factors N and GD. *J. Abn. & Soc. Psychol.*, 1939, **34**, 239–248.
6. Guilford, J. P., and Martin, H. G. *The Guilford-Martin personnel inventory.* Beverly Hills, California: Sheridan Supply Company, 1943.
7. Mosier, C. I. A factor analysis of certain neurotic tendencies. *Psychometrika*, 1937, **2**, 263–287.
8. Terman, L. M., and Miles, C. C. *Sex and temperament: studies in masculinity and femininity.* New York: McGraw-Hill Book Company, 1936.

Measuring Progress in Radio Training

Gordon L. Macdonald

New York University

In New York City the War Department in conjunction with the New York State Department of Education set up a school in 1942 to train radio technicians where the instruction was administered by various universities and radio schools. The course of instruction, lasting six months, was divided into two sections of three months each. It was within this educational framework that the problem for this study evolved.

This study grew out of the need for an adequate test to measure the progress of trainees under instruction. Specifically the problem is twofold: (1) to develop an adequate instrument for measuring progress in training that would be valid and reliable at various levels; and (2) to apply the instrument to groups of trainees for the purpose of evaluating their progress and comparing results in the several institutions.

This investigation included 1,156 subjects studying the theory and practice of radio servicing (how to repair radios). These subjects represent three levels of general education: (1) High School Students, consisting of three groups of 28, 34 and 25 subjects between the ages of 16 and 20 studying in a public vocational high school (These students were not in any way involved in the War Department training program and were included in this study for comparative purposes); (2) High School Graduates, consisting of four adult groups of 126, 289, 326, and 216 students between the ages of 18 and 40. This age limit was set by the War Department for radio trainees accepted for a course of study in a commercial radio school under the supervision of the War Department and the New York State Department of Education; and (3) College Students of which there were three groups of 25, 30, and 34, with a mean age of 20 years, studying at three universities in New York State where their courses were limited to three months.

The subjects comprising the two educational levels of adult High School Graduates and College Students, who were trainees of the Signal Corps, War Department, were chosen on the basis of a Civil Service aptitude test and a physical examination by army doctors, and they were further required to join the enlisted reserve of the Army of the United States. The group of College Students were further selected on the basis of their knowledge of higher mathematics and college physics.

There was no standardized objective achievement test in the field of radio available. However, an unstandardized test was discovered consisting of two forms, A and B, each containing 116 items and it was used and standardized in this investigation. It was a paper and pencil, multiple choice, self administering, group achievement test with no time limit, but requiring from one hour and a half to over two hours for completion.

The preliminary standardization was done at the level of High School Graduates of general education, and at two levels of specialized radio training. Group V with three months' training, and Group VII with six months' training were used in this phase of the study.

As a result of these analyses two new forms of the test, revised Form A and Form B, were prepared consisting of the best 99 items from the 116 of each original form. It is these forms which were used in the rest of this investigation.

Training Progress Measured

The revised forms were administered to various groups at three distinct levels of general education to find out if progress in specialized radio training could be measured equally well with the test regardless of background. Norms in radio training thereby can be made available taking into account any contribution of general education.

High School Students (Groups I, II, and III): The subjects at this level of general education had completed the 2½ years of general instruction of a public vocational high school and were in attendance on one of the last three terms of specialized radio instruction. They were tested at the end of 315 hours of radio instruction, or one term (Group I), at the end of 630 hours of radio instruction, or two terms (Group II), and at the end of 945 hours of radio instruction, or graduation (Group III). Both Forms A and B, with some exceptions, were administered to all members of these groups.

The means, sigmas of distribution, and C.R.'s between the means on both forms, for the three groups, are given in Table 1.

The levels of specialized radio training are differentiated one from the other by both forms of the test as is indicated by the difference between the means; and these differences are significant as indicated by the critical ratios.

High School Graduates (Groups IV, V, VI, and VII): The adult radio trainees who comprised this level of general education were all high school graduates. The groups were divided according to specialized training in radio as follows: Group IV being novices with no training; Groups V and VI being students with three months' full-time training of 624 hours; and Group VII having completed the six months' full-time course of training of 1,248 hours.

Table 1

Progress in Specialized Radio Training at the General Educational Level of High School Students, Showing M, σ (dist.), for Each Level, and C.R.'s between Levels of Specialized Training Separately for Forms A and B

Form A				
Training Period	Group	Mean	σ (dist.)	N
One term (315 hours)	I	43.00	7.77	28
Two terms (630 hours)	II	49.80	11.61	34
Three terms (945 hours)	III	60.92	10.52	25
Critical Ratios				
One term (315 hours) vs. two terms (630 hours)				2.75
One term (315 hours) vs. three terms (945 hours)				7.00
Two terms (630 hours) vs. three terms (945 hours)				3.81
Form B				
Training Period	Group	Mean	σ (dist.)	N
One term (315 hours)	I	39.00	5.70	28
Two terms (630 hours)	II	46.60	10.17	34
Three terms (945 hours)	III	61.95	9.03	25
Critical Ratios				
One term (315 hours) vs. two terms (630 hours)				3.70
One term (315 hours) vs. three terms (945 hours)				10.93
Two terms (630 hours) vs. three terms (945 hours)				6.10

The means, sigmas of distribution, and C.R.'s between the means for all groups on both forms are given in Table 2.

It is evident from the means given in Table 2 that the three levels of specialized radio training of "no training," "three months' training," and "six months' training" are distinguished one from the other. The critical ratios indicate that the differences between the means of the two groups with three months' training (Groups V and VI) are statistically insignificant, while the differences between the means for the other levels of specialized training are highly significant, these values holding for both forms of the test.

The results with Form A (Table 2) show that the difference between the means of the two groups tested at the end of three months of specialized radio training (Groups V and VI) is statistically insignificant, the C.R. being 1.36. The C.R. between the three months' training level (Group V) and the six months' training level (Group VII) is 10.57, and

the C.R. between the three months' training level (Group VI) and the six months' training level (Group VII) is 9.39. Thus Form A distinguishes clearly between levels of specialized training in radio among high school graduates.

The results for Form B are similar. The novices, or those with no training (Group IV) are distinguished from the others quite significantly with C.R.'s of 14.08 and 14.74 in comparison with the two groups with three months' training (Groups V and VI); and with a C.R. of 20.12 between them and those at the six months' training level (Group VII). The C.R. of the means of the two groups with three months' training (Groups V and VI) is statistically insignificant, as on Form A, while the C.R.'s of the means between these same groups at the three months' level

Table 2

Progress in Specialized Radio Training at the General Educational Level of High School Graduation, Showing M, σ (dist.), for Each Level, and C.R.'s between Levels of Specialized Training Separately for Forms A and B

Form A

Training Period	Group	Mean	σ (dist.)	N
Three months	V	48.80	9.66	144
Three months	VI	50.35	10.32	162
Six months	VII	62.65	13.43	107

Critical Ratios

Three months (624 hours) Group V	vs.	Three months (624 hours) Group VI	1.36
Three months (624 hours) Group V	vs.	Six months (1,248 hours) Group VII	10.57
Three months (624 hours) Group VI	vs.	Six months (1,248 hours) Group VII	9.39

Form B

Training Period	Group	Mean	σ (dist.)	N
No training	IV	27.30	12.25	126
Three months	V	46.60	9.92	145
Three months	VI	47.35	10.33	164
Six months	VII	60.30	12.77	109

Critical Ratios

No training Group IV	vs.	Three months (624 hours) Group V	14.08
No training Group IV	vs.	Three months (624 hours) Group VI	14.74
No training Group IV	vs.	Six months (1,248 hours) Group VII	20.12
Three months (624 hours) Group V	vs.	Three months (624 hours) Group VI	.64
Three months (624 hours) Group V	vs.	Six months (1,248 hours) Group VII	9.26
Three months (624 hours) Group VI	vs.	Six months (1,248 hours) Group VII	8.87

Table 3

Progress in Specialized Radio Training at the General Educational Level of College Students, Showing M, σ (dist.), for Each Level, and C.R.'s between Levels of Specialized Training Separately for Forms A and B

Form A

Training Period	Group	Mean	σ (dist.)	N
Two months (416 hours)	IX	65.12	6.63	28
Three months (624 hours)	VIII	74.12	9.39	25
Three months (624 hours)	IX	75.00	9.51	30
Three months (624 hours)	X	76.09	12.00	34

Critical Ratios

Two months (416 hours) Group IX	vs.	Three months (624 hours) Group VIII	3.27
Two months (416 hours) Group IX	vs.	Three months (624 hours) Group IX	4.18
Two months (416 hours) Group IX	vs.	Three months (624 hours) Group X	3.92
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group IX	.54
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group X	.71
Three months (624 hours) Group IX	vs.	Three months (624 hours) Group X	.22

Form B

Training Period	Group	Mean	σ (dist.)	N
One month (208 hours)	IX	31.34	8.13	28
Three months (624 hours)	VIII	69.56	8.73	25
Three months (624 hours)	IX	75.10	10.05	30
Three months (624 hours)	X	73.54	12.96	34

Critical Ratios

One month (208 hours) Group IX	vs.	Three months (624 hours) Group VIII	16.48
One month (208 hours) Group IX	vs.	Three months (624 hours) Group IX	18.38
One month (208 hours) Group IX	vs.	Three months (624 hours) Group X	15.63
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group IX	2.05
Three months (624 hours) Group VIII	vs.	Three months (624 hours) Group X	1.40
Three months (624 hours) Group IX	vs.	Three months (624 hours) Group X	.54

of training and that of the group with six months' training (Group VII) are statistically significant, being respectively 9.26 and 8.87. Thus specialized training levels are clearly differentiated by both forms at the general level of education represented by High School Graduates.

College Students (Groups VIII, IX, and X): The radio trainees comprising this level of general education received their specialized radio training at three universities in New York State, the course lasting for three months or 624 hours of instruction. The groups are distinguished from each other according to the university which they attended. Group

IX was tested at the end of one month's training with Form B, at the end of two months' training with Form A, and at the end of three months' training with both Form A and B. Groups VIII and X were tested at the end of three months' training with both forms.

The means and sigmas of distribution of both forms at the different levels of specialized training, and the C.R.'s indicating the statistical significance of the difference between levels of training are given in Table 3.

The means indicate that the levels of specialized training at one, two, and three months are differentiated from each other. The differences on

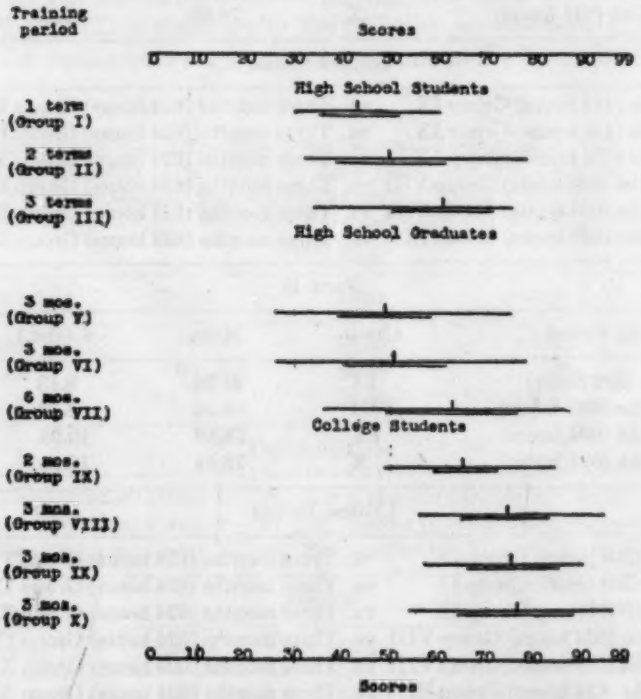
Training period
Scores
0 10 20 30 40 50 60 70 80 90 99
High School Students
1 term (Group I)
2 terms (Group II)
3 terms (Group III)
High School Graduates
3 mos. (Group V)
3 mos. (Group VI)
6 mos. (Group VII)
College Students
2 mos. (Group IX)
3 mos. (Group VIII)
3 mos. (Group IX)
3 mos. (Group X)
0 10 20 30 40 50 60 70 80 90 99
Scores

FIG. 1. Levels of radio training, showing M, one σ, and range in scores on Form A for various amounts of radio training classified according to levels of general education.

Form A at two and three months' training are statistically significant: a C.R. of 3.27 existing between Group IX and Group VIII; one of 4.18 for Group IX; and one of 3.92 between Group IX and Group X.

The data for Form B show that the differences between means at one month's training (Group IX) and three months' training (Groups VIII, IX, and X) are highly significant, the C.R. being 16.48 between Group IX and Group VIII; 18.38 for Group IX; and 15.63 between Group IX and Group X.

The C.R.'s of the means for the groups compared at the three months' level of training are statistically insignificant in all cases. This statistical insignificance of the differences between the means coupled above with the evidence of reliable differences between different levels of specialized radio training, indicates that the test is measuring progress in training.

Progress in radio training is being measured within different levels of general education. Differences between levels of specialized training hold for both forms of the test. A graphic representation of these results is given in Figures 1 and 2.

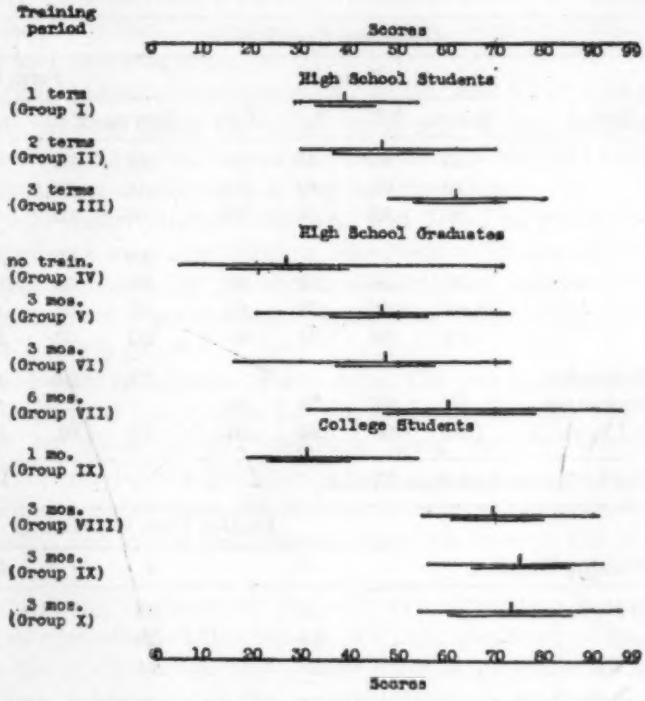
Training period
Scores
0 10 20 30 40 50 60 70 80 90 99
High School Students
1 term (Group I)
2 terms (Group II)
3 terms (Group III)
High School Graduates
no train. (Group IV)
3 mos. (Group V)
3 mos. (Group VI)
6 mos. (Group VII)
College Students
1 mo. (Group IX)
3 mos. (Group VIII)
3 mos. (Group IX)
3 mos. (Group X)
0 10 20 30 40 50 60 70 80 90 99
Scores

FIG. 2. Levels of radio training, showing M, one σ, and range in scores on Form B for various amounts of radio training classified according to levels of general education.

Analysis of Reliability and Validity

The measurement of progress having been indicated the next step was to check the equality of the forms and verify the reliability at the various levels of specialized radio training, and the different levels of general education; and to establish the validity of the test.

Equality of the Forms. The two forms were compared at the different levels of specialized radio training within the levels of general education. Both forms were found to be equal in difficulty, with variability of per-

formance being much the same for both forms, with the exception of Group I at the general education level of High School Students, performance on Form A being slightly more variable.[1]

Reliability. The odd-even correlations, corrected by the Brown-Spearman formula for length, for all groups tested are given in Table 4. The parallel form correlations for those groups which were tested with both forms is given in the same table.

Table 4

Odd-Even and Parallel Form Correlation at All Levels of General Education

Odd-Even Correlations								
	Form A				Form B			
Training Period	Raw r_{oe}	P.E.	r_{oe}*	P.E.	Raw r_{oe}	P.E.	r_{oe}*	P.E.
Group I	.53	.09	.69	.07	.53	.09	.69	.07
Group II	.82	.04	.90	.02	.51	.09	.67	.07
Group III	.80	.04	.88	.03	.88	.03	.93	.02
Group IV					.80	.02	.89	.01
Group V	.70	.03	.83	.02	.69	.03	.81	.02
Group VI	.63	.03	.78	.02	.67	.03	.80	.02
Group VII	.84	.02	.91	.01	.82	.02	.90	.01
Group IX (1 month)					.74	.04	.85	.03
Group IX (2 months)	.54	.06	.70	.05				
Groups VIII, IX, and X	.86	.01	.93	.01	.89	.01	.94	.01

* Corrected by Brown-Spearman formula.

Parallel Form Correlations			
Training Period	*N*	*r*	P.E.
Group I	28	.39	.11
Group II	34	.73	.05
Group III	25	.44	.11
Group VIII	25	.87	.03
Group IX	30	.79	.05
Group X	34	.92	.02
Groups VIII, IX, and X	89	.88	.02

High School Students (Groups I, II, and III): The groups at the lowest level of high school training (Group I) have an odd-even *r* of .69 on both forms. This reliability is unsatisfactory at this lowest level of specialized training and level of general education.

[1] In an effort to save space the tables for this analysis have been omitted, but they may be obtained from the author.

The parallel form correlation of .39 for this group is likewise quite low, and is not four times its P.E. Thus any prediction of individual performance on one form as compared to individual performance on the other would likely be no better than chance for members of this group.

The group at the next level of training (Group II) has a reliability coefficient of .90 that is satisfactory on one form while on the other it is much lower (.67) and unsatisfactory. The parallel form correlation is, however, high enough to be acceptable.

The group at the highest level of radio training among high school students (Group III) has reliability coefficients that are acceptable on both forms of the test. However, the parallel form correlation of .44 is quite low and just four times its P.E.

High School Graduates (Groups IV, V, VI, and VII): The reliability coefficients on both forms for all of the groups at the various levels of specialized training indicate that the reliability of the two forms is generally satisfactory at this level of general education.

College Students (Group VIII, IX, and X): The groups with three months' training were combined on one Pearson Product-moment chart for purposes of computing the odd-even correlation because it was shown that their training was apparently identical as indicated by the means of all groups on both forms.

The odd-even correlation (Form A) at the two months' level of specialized training is lower than for the other levels, with the reliability coefficient for three months' training being most satisfactory. The parallel form correlations for the various groups at the three months' level of specialized training are quite satisfactory and indicate that prediction of individual, and group performance from one form to the other would be highly reliable.

The reliability as shown by the odd-even correlation is acceptable at all levels of specialized radio training, with the exception of the first term students (Group I) at the high school level of general education. The parallel form correlation at the general education level of High School Students is not as reliable as at the general education level of College Students where the coefficients are acceptable for group and individual prediction.

Validity. The progress made in specialized radio training was analyzed at three levels of general education, and within those levels groups were differentiated one from the other by the means according to progress in specialized radio training. Those differences between the means for the various levels of specialized radio training within each general educational level was shown to be statistically significant, and was statistically insignificant between groups with specialized radio training. This dis-

tinguishing between levels of specialized radio training indicates a valid test of radio achievement.

Another indication of validity is to be found in the low correlations which were found to exist between the radio achievement test and the Otis Intelligence test. Groups VIII and X were administered the Otis Intelligence Test, Higher Examination Form B, timed for thirty minutes and scored for the number right, and the results correlated with both forms of the radio achievement test. Group VIII has a correlation of $-.03$ between Form A of the radio achievement test and the Otis Intelligence Test and .26 between Form B and the Otis Test; and for Group X a correlation of .05 was obtained between Form A and the Otis Test and .02 between Form B and the Otis Test. Thus we have indicated by the low correlations, approaching zero in three instances, that there is being measured by the radio achievement test some capacity that is different from that defined capacity measured by this standardized and valid intelligence test.

There being no known acceptable outside criteria with which this test may be correlated, the evidence of parallel form correlation is offered in its stead. Each form of the radio achievement test at the general education level of College Students has a high self correlation, and each form correlates high with the other form. Both forms of the radio achievement test may be said to be valid radio achievement tests and they are measuring equally well that capacity which, for this study, is known as radio knowledge.

Summary and Conclusions

The purpose of this investigation was to measure training progress of radio learners, and to standardize the instrument thus used.

Three levels of general education were represented in the subjects selected for measurement of radio training: High School Students, High School Graduates, and College Students. The subjects at the two highest levels of general education were selected from male applicants on the basis of a Civil Service Aptitude test and a physical examination by Army doctors.

A preliminary standardization of two forms of a radio achievement test was made; the two forms compared for difficulty and variability of performance; odd-even correlation computed for reliability, and a bi-serial correlation of test items computed for the two forms. On the basis of this work the test was revised and the forms reevaluated. Both forms were found to be close to equality of difficulty and variability of performance.

The two forms of the test were administered to each member of three groups, with one, two, and three terms of radio training, at the general educational level of High School Students. At the general educational

level of High School Graduates one form (Form B) was administered to a group of novices, and both forms were administered in alternate fashion to the members of the other groups. At the general educational level of College Students the two forms of the test were administered to all members of the three groups; while one group (Group IX) was tested with Form B at the end of one month, and Form A at the end of two months of radio training.

Answers to the following questions were sought from the data thus accumulated:

1. Are both forms of the test reliable at various levels of radio training within the several levels of general education?
2. Is this a valid test of radio achievement?
3. Are the two forms equal in difficulty and variability of performance for various levels of specalized radio training within each general educational level?
4. What prediction of an individual's or a group's progress in training exists for either form?

The answers to these questions are the conclusions derived from this study.

1. The test is a reliable test as is indicated by the odd-even and parallel form correlations shown in Table 4. The odd-even reliability coefficients range from .69 to .93 on Form A and from .67 to .94 for Form B at the various levels of specialized radio training with a median coefficient of .88 for Form A and .85 for Form B. The lowest reliability coefficients were at two levels of specialized radio training within the general educational level of High School Students where the odd-even correlation was .69 for the lowest group (Group I) on Form A and .67 for the middle group (Group II) on Form B. The parallel form correlation for Group I was .39, and .44 for Group III. It appears from this that the test is less reliable at the lowest of the three levels of general education.

2. The test appears to be a valid test of radio achievement at various levels of specialized radio training and general education. The magnitude of the parallel form correlations suggests this as shown in Table 4. The low correlations between both forms of the test and an intelligence test is evidence that the test is measuring some factor other than intelligence, which is assumed to be radio knowledge as that was what the students were studying, and that was what was tested. Essentially however, the test is valid because it differentiates between levels of specialized radio training within each of three levels of general education, and with differences between the mean levels of specialized training that are practically without exception statistically significant as shown in Tables 1, 2, and 3.

3. The two forms are close to equality of difficulty at all the higher levels of specialized radio training for all general educational levels, with Form B slightly more difficult at the lowest levels of specialized training.

4. The coefficients of correlation between levels of specialized radio training at the general educational level of College Students (Group IX) were of such magnitude (.62 to .82) as to be acceptable for group predictions of training progress; and individual prediction as indicated by these same coefficients is increased from 20% to 40% above chance at the various levels of specialized radio training.

In general, it appears that specialized radio training is affected somewhat by the level of general education (see Figures 1 and 2).

Thus it may be concluded that training progress of radio trainees has been measured and a test has been standardized to accomplish it.

Received August 7, 1944.

A Study of the Effect of Music Distraction on Reading Efficiency

Mack T. Henderson, Anne Crews, and Joan Barlow

Grinnell College, Iowa

Very often college students claim that they can study effectively with the radio on, that music does not "bother" them. This suggested the following study which attempts to determine whether or not reading efficiency is influenced when music is used as distraction, and whether there is any difference in the influence of popular and classical music upon reading efficiency. In a similar study Paul Fendrick [1] found that semi-classical music tended to reduce efficiency, but since he did not equate his groups and because he used only one type of distraction, it was decided to supplement his results through this experiment.

Procedure

Fifty freshman women helped us with this experiment. These women were divided into three equally matched groups on the basis of their psychological examination scores (American Council on Education Psychological Examination, 1942 edition) and reading test scores (Nelson-Denny Reading Test) obtained in September, 1943. The A.C.E. means for the No Distraction, Classical, and Popular groups were 107.7 ± 22.6, 102.5 ± 30.3, 103.6 ± 23.8. The means for the vocabulary section of the reading test were 43.1 ± 16.1, 42.6 ± 16.7, 43.8 ± 13.5, and for the paragraph section of the reading test 45.3 ± 11.5, 47.3 ± 11.7, 48.0 ± 13.9. The differences in these means were found to be statistically insignificant as measured by Fisher *t*.

First of all, the subjects filled out a questionnaire which was constructed primarily for the purpose of determining whether or not the subjects were accustomed to studying with the radio on; whether or not they thought that the radio reduced their study efficiency; the amount of studying done with the radio on; and the type of program they usually listened to when studying.

The Nelson-Denny Reading Test was used in this study to measure the reading efficiency of the three groups. This test was chosen for four reasons: (1) it has two forms which made possible the use of one form as

[1] Fendrick, P. The influence of music distraction upon reading efficiency. *J. educ. Res.*, 1937, 31, 264–271.

a pre-test and the other as a final test; (2) Form A had been given to the freshmen when they entered college, September, 1943, making pre-test scores immediately available; (3) the test has two sections, a vocabulary section and a paragraph comprehension section, making it possible to determine the influence of distraction upon these subdivisions as well as upon the total scores; (4) the test required only 30 minutes to complete, 10 minutes for vocabulary and 20 for paragraph comprehension.

Form B of the Nelson-Denny Reading Test was administered, as the final test, to a group of 14 freshman women with popular music as distraction and to a group of 17 with classical music as distraction, while a third group of 19, a control group, took the reading examination without any distraction. Hereafter, these groups will be referred to as Popular, Classical, and No Distraction groups. Popular and classical music were the two types of music chosen for this study, because the questionnaire results showed that they were the two types to which most of the subjects usually listened. Typical, familiar recordings of both types of music were carefully selected to be played during the tests. The recordings used in this experiment are as follows:

Musical Recordings: Popular music (order of presentation): 1. Two O'clock Jump (Harry James); 2. That's What You Think (Krupa); 3. Sunday, Monday, or Always (Frank Sinatra); 4. Mr. Five by Five (Harry James); 5. Prince Charming (Harry James); 6. Tuxedo Junction (Glenn Miller); 7. Idaho (Benny Goodman); 8. Crosstown (Glenn Miller); and 9. Close to You (Frank Sinatra), and *Classical music:* Symphony in D Minor by César Franck (Philadelphia Symphony Orchestra, Victor Recording, 6726–6730).

The conditions under which the tests were given were regulated as carefully as possible. The tests were administered on three successive afternoons at hours when the greatest number of the subjects would be free. However, it was impossible to find times when they were all free; hence, the rather small number of subjects. During the test, the subjects were asked to assume that they were in their own rooms studying with the radio on. The volume of the phonograph was predetermined by a group of judges, including students, and regulated to approximately the same loudness which the subjects ordinarily maintained when studying with the radio as background. These judges, who were stationed at various places in the room in which the tests were to be given, agreed upon the loudness desired. In order to assure some measure of similarity of volume, the position of the volume control was noted and used throughout the experiment. This method was resorted to, because a physical device for determining volume in decibels was not practical, since the volume within each record varied noticeably.

The significance of the differences in the means was determined by the Fisher *t* test.

Results

In Table 1 are recorded the averages of the No Distraction, Classical, and Popular groups, the differences between the averages of the pre-test scores and the final test scores of each group, and the significance of these differences in averages. It will be observed that the only score influenced by the distraction more than could be accounted for by chance is the

Table 1

Nelson-Denny Averages and *t* Scores

	N	Pre-Test (Form A)	Final Test (Form B)	Difference	*t* (Fisher)	P
No Distraction	19					
vocabulary		43.1	50.0	+6.9	1.260	.20
paragraph		45.3	49.2	+3.9	.923	.35
Classical	17					
vocabulary		42.6	48.4	+5.8	.906	.35
paragraph		47.3	46.1	−1.2	.266	.80
Popular	14					
vocabulary		43.8	47.8	+4.0	.605	.55
paragraph		48.0	22.9	−25.1	6.160	<.001

paragraph score of the Popular group. This score was reduced 25.1 score points, on the average, below the pre-test score. It is interesting to note that the vocabulary scores of all three groups showed an increase even though the increases are not statistically significant as measured by Fisher *t*.

In order to determine whether or not differences exist between those who are accustomed to studying with the radio and those who are unaccustomed to studying with the radio, the data of Tables 2 and 3 are presented. These data show that, regardless of the students' study habits, the groups function alike. The paragraph scores of the Popular group showed a significant decrease in the final test score whether students were accustomed to studying with the radio or not; all other test score changes were within the range expected by chance.

In trying to account for the lack of influence or distraction of classical music upon the test results and the lack of influence of popular music upon the vocabulary scores, one can only suggest explanations. A reasonable explanation for the lack of distraction of classical music is that the rhythms and melodies of classical music are usually more complex and less obvious

Table 2

Nelson-Denny Averages and *t* Scores of Those Who Use the Radio When Studying

	N	Pre-Test (Form A)	Final Test (Form B)	Difference	*t* (Fisher)	P
No Distraction	14					
vocabulary		42.3	49.8	+7.5	1.089	.30
paragraph		44.1	48.9	+4.8	1.062	.30
Classical	9					
vocabulary		41.4	50.4	+9.0	1.341	.20
paragraph		47.8	46.9	−.9	.134	.85
Popular	8					
vocabulary		45.5	50.5	+5.0	.670	.50
paragraph		53.8	25.1	−28.6	5.485	<.001

Table 3

Nelson-Denny Averages and *t* Scores of Those Who Do Not Use the Radio When Studying

	N	Pre-Test (Form A)	Final Test (Form B)	Difference	*t* (Fisher)	P
No Distraction	5					
vocabulary		36.8	40.6	+3.8	.531	.60
paragraph		39.6	40.0	+.4	.042	.95
Classical	8					
vocabulary		43.3	46.1	+2.8	.245	.80
paragraph		46.8	45.3	−1.5	.201	.85
Popular	6					
vocabulary		41.5	44.2	+2.7	.244	.80
paragraph		40.3	20.0	−19.7	3.849	<.001

than those of popular music. The simpler and obvious rhythms and melodies of popular music are easily grasped by a group of subjects and are therefore listened to by the subjects. Naturally, while they listen to the music their attention is diverted from the task at hand. Classical music with its subtle rhythms and hidden melodies is apt to be vague and is therefore not "listened to." It becomes a background against which the assigned task is accomplished without interference, and under these conditions it does not divert the subject's attention from his work. Just what a group of persons highly trained in the understanding of classical music would do under the conditions of this experiment remains to be determined.

A likely explanation for the fact that popular music influences the paragraph scores and not the vocabulary scores seems to lie in the nature of the test materials. The paragraph materials are meaningfully related and require sustained effort on the part of the subject. In contrast to this, the vocabulary materials are intermittent and unrelated. This suggests that popular music interfered with the more complex of the two test sections.

The suggested explanations might be summarized by saying that whether or not music is a real distraction depends upon the complexity of the music and upon the complexity of the test materials. In this experiment, the subtler music (classical) did not influence the test results, and the obvious music (popular) influenced only the paragraph section of the test.

Conclusions

1. Popular music distracted a group of subjects significantly on the paragraph section of the Nelson-Denny Reading Test. Classical music showed no evidence of distraction in either the vocabulary or paragraph sections of the test, nor did the popular music show evidence of distraction upon vocabulary.

2. Students accustomed to studying with the radio were influenced as much or as little as students unaccustomed to studying with the radio.

3. It is suggested that whether or not music serves as a distraction depends upon the complexity of the music and upon the complexity of the test materials.

Received June 2, 1944.

Book Reviews

Note: The length of a review of a book or monograph is no indication of its importance. Because of W.P.B. restrictions on paper reviewers of books have been requested to prepare their reviews in fewest possible words. Ed.

Halsey, George D. *Making and using industrial service ratings.* New York: Harper & Bros., 1944. Pp. xxii + 149. $2.50.

"This book is intended primarily for operating and personnel executives who are interested in the practical aspects of service ratings—what has been done, what the difficulties are, and how rating problems similar to theirs have been worked out successfully by other executives." The book, however, contains material of value to anyone interested in employee ratings.

Ratings of nonsupervisory, supervisory and executive personnel are discussed. Numerous helpful excerpts and sample forms are presented, mostly from the field of public personnel administration.

The author gives proper emphasis to such practical points as how to train supervisors to rate employees and to discuss the ratings with the individuals.

"Service ratings have two distinct and different purposes: to serve as an aid in training and supervision, and to furnish an evaluation of the person's job performance as an aid in making sound administrative decisions—salary increases, layoffs, etc." In developing a rating plan it is important to distinguish carefully between these two purposes. The rating procedure will differ in many respects depending on whether it is designed to serve one or both of these purposes.

The efficiency rating system of the federal government is explained in considerable detail. It is evident that this plan is superior in every way to the graphic rating scale used almost universally in industry.

With regard to the problem of scoring a rating form, the approach is both laborious and naive. From an attempt to attach numerical grades to employee ratings problems arise which cannot be overcome in a practical situation. Logical and statistical considerations, which the author completely overlooks, prevent one from obtaining a total score "by adding the numerical values of the ratings on the separate qualities." There is a wide-spread misconception on this point among industrial people. It is not clear why it is necessary to grade the ratings in order to serve the training or administrative purposes outlined above.

It is unfortunate that the book does not touch on the problem of using employee ratings in a unionized organization. The broader problem is how can the merit of certain individuals be appraised and rewarded under conditions of collective bargaining.

This book is a worthwhile contribution to the literature on rating industrial employees. Although the approach is not scholarly, most of the content is sound and, above all, it is realistic.

Charles C. Gibbons

Owens-Illinois Glass Co.,
Toledo, Ohio

Cason, Eloise Booker. *Mechanical methods for increasing the speed of reading. An Experimental Study at the Third Grade Level.* New York: Columbia University Contributions to Education, No. 878, 1943. Pp. 80. $1.75.

This monograph is a complete report of an experiment set up to answer a specific question for a specific group of subjects: Does intensive training in correct eye movements improve measurable reading skills to a greater extent than would the use of the same amount of time for free library reading? The experiment is admirably designed to produce an unequivocal answer. Paired groups of third grade children (25 pairs in School A, 26 pairs in School B) were used as subjects. Pairing of individuals was based on scores from the Gates Reading Survey Test, but Tables 1 and 2 indicate that the groups were closely similar on the Otis Intelligence Test as well. Twenty-minute training periods were scheduled five times a week for four weeks under the direction of the regular teachers, no other reading instruction being given during the experiment. In School A, experimental subjects read printed material marked and spaced to call attention to phrases. Control subjects read library materials in the same room. In School B, experimental subjects practiced with the Metron-O-Scope. Control subjects read library materials in another room. Before and after the four-week training period, the Gates Reading Survey Tests and several new tests specially constructed to measure skill in the reading of phrases were given. Results bearing on the main problem were definite and clear-cut. *In neither school was there a statistically significant difference on any test between the means of the experimental and control groups.*

The data were then analyzed in several other ways. In order to throw light on the question as to whether the effects of method may differ for children of different ability levels, each group was divided into upper, middle, and lower thirds. Since this would place only about eight or nine persons in each subgroup one would anticipate difficulty in demonstrating that any differences in gains were statistically significant. Therefore, it is interesting to find, using the t-test, that a few were. In the School A Phrase Material Group, the upper third lost in speed while the middle third gained. In the Metron-O-Scope Group the upper third gained significantly more in comprehension than the lower third. In the School B Library Group, the middle third gained significantly more in comprehension than the lower third. Though results based on so few cases cannot be conclusive, they suggest the desirability of special investigation of this problem. Another sort of analysis used the results of a re-administration of the Gates Test after summer vacation. There were still no reliable differences between equated groups. It seemed to this reviewer that the most interesting difference in the whole study was one which the author declined to interpret because of uncontrolled factors in the two situations, namely the difference between schools in gains on the Gates Test. In School A there were no significant gains for either group. In School B, both experimental and control groups showed gains of one-half to three-fourths of a year of reading age, gains which persisted over the summer vacation period. When one remembers that the practice periods added up to less than seven hours altogether, these figures seem very large. Whatever the uncontrolled factors were which accounted for this result in School B may well be of more practical importance than the factors the experiment was set up to test.

There are a few places where the statistical explanations are a little obscure, but there seem to be no errors in procedure which would cast doubt on the dependability of the results. The findings of the study as a whole will lend support to the arguments of those who hold that mechanical and motor skills are less important aspects of reading than are intellectual and motivational factors. Since free reading is much simpler and less expensive than the mechanical methods, the results should have definite practical application.

Leona E. Tyler

University of Oregon

Sargent, S. Stansfeld. *The basic teachings of the great psychologists.* New York; New Home Library, 1944. Pp. xiv + 346. $.69.

The book covers about all the major topics in psychology with a pretty fair balance. It includes most of the applied field as well as the general field, but the former is usually somewhat implicit in the discussions of the latter rather than being set off by itself as applied psychology. In each chapter the treatment is essentially chronological, comprising a brief summary of the findings or theories of quite a number of people on the topic under consideration. The treatment is not critical, and the controversial questions are left open.

The author indicates that it is written for the layman and secondarily perhaps as a refresher for advanced students. The reviewer believes that these two markets should be reversed in importance. The layman will find the discussion somewhat cluttered by the names of so many people unfamiliar to him, and he will not find an extensive topic in which he might be interested treated at any length. It is, of course, impossible to cover so much ground in any very systematic fashion. Occasionally, too, some technical term is introduced without any explanation, although later in the book it may be discussed in more detail. On the other hand, the advanced student might find it quite helpful as a sort of thumbnail review of the whole field, and he would be almost certain to encounter some bits of information here and there that would be a helpful supplement to his existing knowledge on a topic. The author must have covered a tremendous amount of source material in order to formulate the concise paragraphs which he does on so many topics by so many people. From that standpoint the job is well done.

The title of the book is subject to criticism. There certainly are not as many "great" psychologists as listed. The biographical notes at the end include some two hundred psychologists. It is probably not the author's fault, but the publisher's blurb on the jacket is in poor taste and smacks of "gold brick."

Harold E. Burtt

Ohio State University

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota

And now to live again. Betsy Barton. New York: D. Appleton-Century, 1944. Pp. 150. $1.75.

Psychology of personnel. Henry Beaumont. New York 3: Longmans, Green & Co., Inc., 1945. Pp. 300. $2.75.

Employment opportunities in characteristic industrial occupations of women. Elizabeth Benham. Bulletin No. 201, Women's Bureau, U. S. Dept. of Labor, 1944. Washington 25, D. C.: U. S. Government Printing Office. Pp. 50. $.10.

Your place at the peace table. Edward L. Bernays. Duell, Sloan & Pearce, Inc., Dept. 3F-12, 270 Madison Avenue, New York 16, N. Y. Pp. 60. $1.00.

Psychology: Principles and applications. T. L. Engle. Yonkers-on-Hudson 5: World Book Co., 1945. Pp. 549. $2.12.

The classification of Jewish immigrants and its implications. A survey of opinion. N. Goldberg, J. Lestchinsky, and M. Weinreich. Yiddish Scientific Institute—YIVO, 535 West 123rd St., New York 27, N. Y., 1945. Pp. 154. $2.00.

The people's choice. Lazarsfeld, Berelson, and Gaudet. New York 16: Duell, Sloan & Pearce, Inc. $3.00.

Braille and talking book reading: A comparative study. Berthold Lowenfeld. American Foundation for the Blind, Inc., 15 W. 16th St., New York 11, N. Y., 1945. Pp. 53. $1.00.

Problems of the postwar world. A symposium edited by Thomas C. T. MacCormick. New York 18: McGraw-Hill Book Co., Inc., 1945.

Human nature and enduring peace. Edited by Gardner Murphy. Third Yearbook of the Society for the Psychological Study of Social Issues. New York: Houghton Mifflin Co., 1945. $3.50.

Jobs for the physically handicapped. Louise Neuschutz. New York: Bernard Ackerman, Inc., 1944. Pp. 204. $3.00.

Selling as a postwar career. David R. Osborne. Chicago: The Dartnell Corp., 1944. Pp. 83. $1.00.

Understanding as a condition for success in order-giving. Paul and Faith Pigors. Industrial Relations Associates, Inc., Cambridge, Mass., 1945. Pp. 28.

Soldier to civilian. G. K. Pratt. New York 18: McGraw-Hill Book Co., Inc., 1945. Pp. 233. $2.50.

The New York Hospital: A history of the psychiatric service, 1771–1936. W. L. Russell. New York 27: Columbia Univ. Press, 1945. Pp. xiv + 594. $7.50.

Intelligence and its deviations. Mandel Sherman. New York 10: The Ronald Press Co., 1945. Pp. 300. $3.75.

Guidance and personnel services. Ruth Strang and Robert Hoppock. Occupational Index, Inc., New York University, New York 3, N. Y. Pp. 6. $.25.

Psychology in education. J. B. Stroud. New York 3: Longmans, Green & Co., 1945.

Techniques of guidance. A. E. Traxler. New York 16: Harper & Bros., 1945. Pp. 381. $3.50.

The job of the industrial counselor for women. Frances W. Triggs. Washington, D. C.: U. S. Office of Education, 1944. Pp. 32. Free.

Elementary statistics for students of education and psychology. E. B. Van Ormer and C. O. Williams. New York 3: Longmans, Green & Co., Inc., 1945. Pp. 120. $1.75.

A study of transfer of training from high school subjects to intelligence. A. G. Wesman. New York: Bureau of Publications, Teachers College, Columbia University, 1945. Pp. 82. $1.75. (Contributions to Education No. 909)

Psychology of teaching. A. D. Woodruff. New York 3: Longmans, Green & Co., Inc., 1945. Pp. 200. $1.75.

Counseling for veterans. Proceedings of Conference January 18–19–20, 1945, sponsored by Minneapolis Vocational Guidance Association, Inc., Minneapolis 14: Center for Continuation Study, University of Minnesota. Pp. 151. $1.00.

Experience in retraining on the Dvorak keyboard. American Management Association, 330 W. 42nd st., New York 18, N. Y. $.75. (Supplementary Office Management Series No. 1)

How to prepare and publish an employee manual: Based on a survey of company practices. New York 18: American Management Association. $.75.

The negro worker: An analysis of management experience and opinion on the employment and integration of the negro in industry. New York 18: American Management Association. $.75.

Supervision of women on production jobs: A study of management's problems and practices in handling female personnel. New York 18: American Management Association. $.75.